APPENDIX P. SEDIMENT QUALITY TRIAD LINES OF EVIDENCE FOR THE BERA OF LPRSA BENTHIC INVERTEBRATES

Table of Contents

Tables

Figures

Acronyms

AFDW	ash-free dry weight
AIC	Akaike's information criterion
ASTM	American Society for Testing and Materials
AVS	acid volatile sulfide
BERA	baseline ecological risk assessment
BHC	benzene hexachloride
BIC	Bayesian information criterion
COI	chemical of interest
COPEC	chemical of potential ecological concern
CPG	Cooperating Parties Group
CSM	conceptual site model
DDD	dichlorodiphenyldichloroethane
DDE	dichlorodiphenyldichloroethylene
DDT	dichlorodiphenyltrichloroethane
DO	dissolved oxygen
EFA	exploratory factor analysis
EqP	equilibrium partitioning
ERA	ecological risk assessment
ERL	effect range-low
ERM	effect range-median
HBI	Hilsenhoff Biotic Index
HPAH	high-molecular weight polycyclic aromatic hydrocarbon
ID	identification
IQR	interquartile range
LOE	line of evidence
LPAH	low-molecular weight polycyclic aromatic hydrocarbon
LPRSA	Lower Passaic River Study Area
LRM	logistic regression model
mERMq	mean effect range-median quotient
MLR	multiple linear regression

mPECq	mean probable effects concentration quotient
MSD	minimum significant difference
NJDEP	New Jersey Department of Environmental Protection
NOAA	National Oceanic and Atmospheric Administration
NYSDEC	New York State Department of Environmental Conservation
OC	organic carbon
PAH	polycyclic aromatic hydrocarbon
PC	principal component
PCB	polychlorinated biphenyl
PCA	principal component analysis
PCDD	polychlorinated dibenzo-*p*-dioxin
PCDF	polychlorinated dibenzofuran
PEC	probable effects concentration
ppt	parts per thousand
PRESS	predicted residual error sum of squares
QA/QC	quality assurance/quality control
REMAP	Regional Environmental Monitoring and Assessment Program
RM	river mile
SDI	Swartz's dominance index
SEM	simultaneously extracted metals
SOP	standard operating procedure
SPI	sediment profile imaging
SQT	sediment quality triad
SVOC	semivolatile organic compound
T20	20% probability of observing toxicity
T50	50% probability of observing toxicity
TCDD	tetrachlorodibenzo-*p*-dioxin
TEQ	toxic equivalent
TOC	total organic carbon
Total DDx	sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT and 4,4′-DDT)
USEPA	US Environmental Protection Agency

USGS	US Geological survey
VIF	variance inflation factor
Windward	Windward Environmental LLC
WOE	weight of evidence

1 Introduction

This appendix to the baseline ecological risk assessment (BERA) for the Lower Passaic River Study Area (LPRSA) provides the methods and results for the three lines of evidence (LOEs) evaluated using the sediment quality triad (SQT) approach. The three LOEs are composed of the benthic invertebrate community in the LPRSA (Section 2), toxicity of LPRSA sediment to benthic invertebrates (Section 3), and sediment chemistry (Section 4). Results from these LOEs are compared using a weight of evidence (WOE) approach, described in the main text of the BERA (Section 6.1). Tables providing data and data summaries related to the LOE and WOE analyses are provided in Appendix B.

The following sections present, by LOE, the methods used to collect and analyze data, a summary of results, and a comparison of LPRSA results to relevant reference area data. The methods were initially developed by the Cooperating Parties Group (CPG) and then modified by US Environmental Protection Agency (USEPA) Region 2 (USEPA 2015a, b).

Quantitative analyses to assess uncertainty in several LOEs are described and presented in LOE-specific uncertainty sections herein. The quantitative analyses are used as a bound on uncertainty for the assessment of risks to benthic invertebrates. This approach is intended to address several significant uncertainties, which are detailed in this appendix and in Section 6.1 of the BERA main text.

2 Benthic Invertebrate Community Line of Evidence

The LPRSA benthic invertebrate community is one of three LOEs evaluated in the SQT approach. Benthic community survey data provide insight into the structure and function of benthic communities as a whole, which is why these data are integral to the SQT assessment. This section describes the methods used to directly assess the LPRSA benthic invertebrate community, the results, and comparison of LPRSA data with reference data.

The LPRSA benthic invertebrate community was described using several metrics calculated from the identified and enumerated taxa at each sediment location; the metrics describe different aspects of the benthic community structure (i.e., abundance [per m^2], richness, diversity, evenness, dominance, and tolerance of environmental stress). The assessment of benthic metrics was based on a comparison of metrics that describe the LPRSA with those calculated for reference locations.

2.1 METHODS

The benthic invertebrate community was surveyed at 97 locations in the LPRSA (Figure 4-5 in the BERA main text) that were co-located with sediment toxicity testing and chemistry analysis sampling locations.[1] Four replicate samples were collected at each location from a sediment depth of 15 cm (6 in.) using a grab sampler and were analyzed independently (rather than composited). Qualified taxonomists identified benthic invertebrates to the lowest practical taxonomic level in three of the replicate samples. The fourth replicate was archived for use in the event that a sample was damaged during shipment or that any other issues developed during data evaluation. Details of the benthic invertebrate community survey methods, quality assurance/quality control (QA/QC) procedures, and results are presented in the *Fall 2009 Benthic Invertebrate Community Survey and Benthic Field Data Collection for the Lower Passaic River Study Area* (Windward 2014).

As agreed upon by USEPA and CPG (Windward 2009), the provisional delineation between the estuarine and freshwater portions of the LPRSA for evaluating the benthic invertebrate community was set at river mile (RM) 8.5 based on the presence or absence of polychaetes in previous benthic invertebrate community surveys conducted by Aqua Survey (2005).[2] Samples from between RM 0 and RM 8.5 were

[1] Sediment was collected for chemical analysis and sediment toxicity testing at one additional location for a total of 98 chemistry and toxicity testing locations. The coarse substrate at this location made it difficult to collect the four replicate benthic invertebrate community samples.

[2] The salinity zones were modified as additional salinity data became available. Based on an in-depth review of the benthic community data from the LPRSA, it was determined that taxa that are more sensitive to changes in salinity are not present in areas of the river affected by daily and seasonal salinity changes. The revised zones used for the assessment of the benthic invertebrate community are outlined in the BERA main text (Section 2.2 and shown in Figure 2-7).

collected and processed using estuarine methods (i.e., collection of the top 15 cm of a 0.1-m^2 area and identification of all organisms retained after field-sieving the sediment through a 1-mm stainless steel screen). Samples from between RM 8.5 and RM 17.4 were collected and processed using freshwater methods (i.e., collection of the top 15 cm of a 0.05-m^2 area and identification of organisms retained after field-sieving the sediment through a 0.5-mm stainless steel screen). At three shallow locations near Dundee Dam (RM 17.4) where sediments were too coarse for the grab sampler (LPRT17B, LPRT17C, LPRT17E), 1-m kick nets were used to sample the benthic community.[3] Sample processing was conducted at a field facility before the preserved samples were shipped to the taxonomy laboratory.

As discussed in Section 2.1.1.1 of the BERA main text, the LPRSA, which is influenced by freshwater input and tides, has a transitional fluvial estuarine zone (between the upper estuarine and tidal freshwater zones) that fluctuates in salinity, both seasonally and daily. During low-flow conditions (such as those experienced during the fall 2009 benthic survey), this fluvial estuarine zone extends from approximately RM 4 to RM 13 and influences the structure of the benthic invertebrate community in the LPRSA. Accordingly, the following salinity zones were used to compare LPRSA benthic invertebrate community data to reference area data: upper estuarine (RM 0 to RM 4), fluvial estuarine (RM 4 to RM 13), and tidal freshwater (RM 13 to RM 17.4). These salinity zones are shown in Figure 2-2 of the BERA main text.

The benthic invertebrate community was quantitatively evaluated using a variety of standardized benthic metrics, which provided a way to condense the lists of taxa into relevant biological information. The metrics used included abundance per m^2, richness (i.e., the number of species [or higher order taxa, if not identified to species] in a sample), the Shannon-Wiener diversity index (Shannon 1948), Pielou's evenness index (Pielou 1966), and Swartz's dominance index (SDI) (Swartz et al. 1985). The Hilsenhoff Biotic Index (HBI) (Hilsenhoff 1987) was also included to describe the freshwater benthic invertebrate community's tolerance of stress. All metrics are presented as the mean of the three replicate samples evaluated at each sampling location.

Individual metrics are useful for describing facets of community structure, and the use of multiple metrics is common when comparing benthic communities (Alden et al. 2002). A summary of each metric used to describe the benthic invertebrate community observed in the LPRSA is provided in the subsections below. Metrics were calculated as the mean metric value based on communities surveyed in the three replicate samples collected at each sampling location.[4] USEPA has recommended that metrics also be calculated by first averaging the abundance of invertebrates across the three replicates, then calculating a single community metric for each sampling location. A

[3] The benthic community data collected using kick nets are discussed qualitatively in the BERA. Those data were not used to characterize risk because of the difficulty in comparing data collected using the different methodologies.

[4] Only two replicates were surveyed at LPRSA SQT location LPRT02E.

discussion of this alternative method for calculating benthic community metrics is presented in Section 2.3.4.

2.1.1 Abundance (per m^2)

Abundance describes the number of benthic invertebrate individuals observed in a sediment sample. Sampling and analytical methods can affect the determination of abundance values, so the total number of individuals collected at each sampling location was corrected to account for subsampling (if less than 100% of a sediment sample was used to determine abundance) and for the surface area sampled. The abundance data reported in this assessment was, therefore, an extrapolated estimate of abundance over the 1-m^2 area. For reporting purposes, it is advantageous to use abundance per m^2 (sometimes referred to as "density") because it is comparable among multiple sampling areas and/or surveys that used different sampling methodologies.

In the LPRSA, samples from between RM 0 and RM 8.5 were collected from a 0.1-m^2 area and processed using a 1-mm sieve, and samples from between RM 8.5 and RM 17.4 were collected from a 0.05-m^2 area and processed using a 0.5-mm sieve. Thus, corrected abundances were divided by the collection area (i.e., either 0.05 or 0.1 m^2) (to achieve a per m^2 unit), depending on the location and the method of collection.[5]

Abundance can increase or decrease in the presence of chemical contaminants (e.g., organic enrichment) or other perturbations (Pearson and Rosenberg 1978). At very high levels of physical and/or chemical disturbance, abundance tends to be reduced as compared with abundance in systems that are relatively unstressed (or have been restored) and that contain pollution-intolerant species. However, under moderately stressed environmental conditions, the benthic community may have a high abundance of tolerant species that can flourish in a moderately disturbed environment. This non-monotonic response of the abundance metric makes it unreliable for assessing responses due to perturbations. Measures of abundance are, therefore, less useful than measures of community structure (Nilsson and Rosenberg 2000; Rosenberg 2001).

2.1.2 Taxa richness

Taxa richness is the total number of taxa present in a sample identified to the lowest practical taxon, and represents a basic measure of biodiversity commonly used to

[5] Reference area samples from Jamaica Bay, Mullica River, and Great Bay were collected from a 0.04-m^2 area (and processed using a 0.5-mm sieve) (USEPA 2001a); accordingly, abundances were multiplied by a factor of 25 to obtain the area-normalized abundance per m^2. Reference samples collected in the Passaic River from the area above Dundee Dam were collected from a 0.05-m^2 area and were adjusted by a factor of 20, similar to freshwater samples from the LPRSA. Reference area samples are discussed in Section 6.1.3.

describe benthic communities.[6] Taxa richness data are presented in this assessment as the mean of the three replicate benthic community samples.

Similar to abundance, taxa richness is not always a straightforward metric for describing communities. Richness can increase relative to stable, mature communities during initial periods of low-intensity disturbance, because the disturbance of stable niches in mature communities allows for competition for a resource among a greater number of taxa (Sousa 1980; Molles 2005; Connell 1978). Given that the LPRSA is a tidal salt-wedge system that receives seasonal freshwater discharge pulses (e.g., during storm events) (Section 2.1 of the BERA main text), benthic communities are expected to be in a state of salinity-driven disturbance. For this reason, taxa richness within the LPRSA has been treated as monotonic throughout the analysis of benthic invertebrate communities; locations with the greatest taxa richness are expected to represent more stable and higher quality habitat (e.g., greater habitat complexity and/or resource availability) for benthic invertebrate taxa.

2.1.3 Diversity

Diversity describes the heterogeneity in a community with respect to the number of taxa (i.e., taxa richness) and the distribution of individuals among those taxa (Gray and Pearson 1982; Peet 1974). In practice, diversity metrics are used to describe the habitat quality of an ecosystem, assuming that a complex and suitable habitat that is relatively unimpacted by anthropogenic stressors allows for a greater diversity of taxa, including both species that are tolerant and intolerant of certain stressors (Heck and Wetstone 1977; Gorham and Alevizon 1989; Butler 1988; Quinn and Peterson 1996; Taniguchi and Tokeshi 2004).

The Shannon-Wiener diversity index (H′) (Shannon 1948) was used as the diversity metric for the LPRSA analysis. This is an index of the diversity of a community that incorporates taxa richness and the distribution of individual abundances among the taxa present (Peet 1974). The lowest possible H′ value is 0, which is calculated for a community with only a single species; H′ increases with increasing taxa richness and evenness.

[6] Under a conventional definition, taxa richness is not truly a measure of diversity because it does not incorporate both richness and evenness (i.e., proportionality of abundance between taxa) (Gray and Pearson 1982; Peet 1974), sometimes referred to as heterogeneity (Simpson 1949). Others have argued that richness is among the simplest measures of "true diversity," of which there are an infinite number of possible calculations (Rényi 1961; Jost 2006).

H′ is calculated as follows:

$$H' = -\sum p_i \ln(p_i)$$ **Equation 2-1**

Where:

H′ = the diversity index
ln = the natural logarithm
i = an index number for each taxon present in a sample
p_i = the number of individuals within a taxon (n_i) divided by the total number of individuals (N) present in the entire sample

The H′ metric is commonly applied to assess benthic communities. However, under certain circumstances, very different community structures can result in similar H′ values; for example, highly rich but dominated communities (discussed in Section 2.1.4) may be assigned H′ values similar to those of highly even communities with low richness. Furthermore, slight (e.g., statistically insignificant) discrepancies in H′ values may relate to substantially different communities as indicated by other measures of diversity (Jost 2006). For this reason, it is important to use H′ in conjunction with other metrics to interpret or compare results among sampling locations.

2.1.4 Dominance and evenness

Dominance and evenness are two related concepts used to describe benthic community structure. Dominance is the degree to which the abundance in a community is disproportionately distributed among all taxa, while evenness is the degree to which the total abundance in a community is evenly distributed among all taxa.

Two measures of dominance and evenness were used to compare LPRSA benthic communities with benthic communities from reference areas: SDI and Pielou's evenness index (J′).

2.1.4.1 Swartz's dominance index

Dominance is defined as the minimum number of taxa whose cumulative abundance equals (or exceeds) 75% of the total number of individuals in a sample (Swartz et al. 1985). To calculate the SDI, all taxa within a sample are first ranked and ordered according to their contribution to the total abundance. Then, the abundances of the most abundant taxa are summed cumulatively until that number equals 75% of the total abundance. The SDI is the minimum number of taxa required to obtain a cumulative abundance greater than or equal to 75% of the total abundance.

2.1.4.2 Pielou's evenness index

Pielou's evenness index (Pielou 1966) (often referred to as Pielou's J′) is a measure of the proportionality of the distribution of the total abundance across all taxa present in

a sample; it is frequently used as a measure of evenness in community structure analysis (Peet 1975). Pielou's J′ is the ratio of the Shannon-Wiener H′ to the maximum possible H′ (i.e., the value of H′ if the total abundance were perfectly distributed among all taxa); mathematically, the maximum H′ equals the logarithm of taxa richness. Pielou's evenness index ranges from 0.0 to 1.0, with greater values indicating a more even distribution of organisms among the taxa present in the sample.

Pielou's J′ is calculated as follows:

$$J' = \frac{H'}{H_{max}} = \frac{H'}{\ln(S)}$$

Equation 1-2

Where:

J′ = the evenness index

H′ = Shannon-Wiener diversity index

H_{max} = the theoretical maximum value for H′ if all species in the sample were equitably distributed (i.e., natural logarithm [ln] of S, where S is the richness of taxa in a sample)

2.1.5 Hilsenhoff Biotic Index

The HBI provides an indication of a freshwater benthic community's tolerance of the enrichment of organic materials and nutrients, as well as pollutants and other stressors (e.g., impoundments) (Hilsenhoff 1987, 1998). The index ranges from 1 to 10, with greater values indicating a greater abundance of stress-tolerant taxa. The index is an average of the tolerance values assigned to each taxa (i.e., 1 for the least stress-intolerant taxa and 10 for the most stress-tolerant taxa) present in a sample, weighted to the abundance of each taxa; only those species with assigned tolerance values are used in the calculation of the index.

Many tolerance values for species common to the Northeast are not reported by Hilsenhoff (1987) but are reported by Bode et al. (2002), who provide values that are specific to species observed in New York State (e.g., non-arthropod invertebrates such as oligochaetes). Arthropod values were taken from the list reported by Hilsenhoff (1987), which was developed for the Midwest. When available, New York tolerance values were used to calculate the HBI for communities in the LPRSA.

The HBI applies to freshwater communities only, and was used to compare LPRSA tidal freshwater benthic community data with benthic community data from freshwater reference locations. The HBI was not calculated for upper or fluvial estuarine benthic communities.

2.2 RESULTS

2.2.1 Grab samples

This section presents the results of the benthic invertebrate community survey conducted in the LPRSA in 2009 (excluding kick net samples). Tables 2-1 through 2-3 provide summaries of the metrics calculated for the three salinity zones (i.e., upper estuarine [RM 0 to RM 4], fluvial estuarine [RM 4 to RM 13], and tidal freshwater [RM 13 to RM 17.4]). Appendix K contains the data presented by location; the complete set of taxa data is presented in the *Fall 2009 Benthic Invertebrate Community Survey and Benthic Field Data Collection Report for the Lower Passaic River Study Area* (Windward 2014).

Table 2-1. Summary of LPRSA benthic community metrics in the benthic upper estuarine salinity zone (RM 0 to RM 4)

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Sample size	25	25	25	25	25
Minimum	127	2.7	0.43	0.30	1.0
Maximum	3,420	16	2.1	0.89	5.3
Mean	1,030	8.0	1.4	0.69	2.9
Standard deviation	859	3.1	0.41	0.16	1.0
5^{th} percentile	172	4.7	0.57	0.41	1.5
10^{th} percentile	272	4.7	0.91	0.46	2.0
25^{th} percentile	523	6	1.1	0.61	2.3
Median	721	7.7	1.4	0.71	3.0
75^{th} percentile	1,220	9.3	1.6	0.80	3.5
90^{th} percentile	2,464	12	1.7	0.85	4.0
95^{th} percentile	2,748	13	2.0	0.88	4.5

Source: Windward (2014)

LPRSA – Lower Passaic River Study Area
RM – river mile
SDI – Swartz's dominance index

Table 2-2. Summary of LPRSA benthic community metrics in the benthic fluvial estuarine salinity zone (RM 4 to RM 13)

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Sample size	54	54	54	54	54
Minimum	147	2.3	0.15	0.10	1.0
Maximum	37,900	16	1.7	0.85	3.3
Mean	6,818	9.0	1.1	0.53	2.0
Standard deviation	8,055	3.6	0.32	0.17	0.59

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
5th percentile	244	3.6	0.51	0.29	1.0
10th percentile	577	4.3	0.71	0.34	1.1
25th percentile	1,793	6.7	0.88	0.46	1.7
Median	4,445	9.0	1.1	0.53	2.0
75th percentile	7,165	11	1.3	0.61	2.3
90th percentile	20,720	14	1.5	0.71	2.7
95th percentile	23,585	15	1.5	0.73	3.0

Source: Windward (2014)

LPRSA – Lower Passaic River Study Area
RM – river mile
SDI – Swartz's dominance index

Table 2-3. Summary of LPRSA benthic community metrics in the benthic tidal freshwater zone (RM 13 to RM 17.4)

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI
Sample size	18	18	18	18	18	18
Minimum	1,990	8.0	1.3	0.51	2.0	8.3
Maximum	30,600	23	2.0	0.73	5.3	9.7
Mean	10,601	16	1.6	0.61	3.3	9.2
Standard deviation	8,441	3.6	0.19	0.06	0.92	0.36
5th percentile	2,050	11	1.4	0.51	2.3	8.5
10th percentile	2,557	12	1.4	0.53	2.3	8.7
25th percentile	4,265	14	1.5	0.58	2.7	9.0
Median	8,565	16	1.6	0.60	3.2	9.2
75th percentile	12,475	18	1.7	0.64	3.7	9.4
90th percentile	23,540	20	1.9	0.69	4.5	9.5
95th percentile	28,050	21	2.0	0.71	5.0	9.7

Source: Windward (2014)

[a] HBI was determined using NYSDEC tolerance values, with the exception of values for arthropods, which were taken from Hilsenhoff (1987).

HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River Study Area
NYSDEC – New York State Department of Environmental Conservation
RM – river mile
SDI – Swartz's dominance index

2.2.2 Kick net samples

The purpose of this section is to provide a brief evaluation of the benthic invertebrate community samples collected in 2009 using kick nets. Kick nets were used at only three LPRSA locations, where grab sampling was not possible due to the coarseness of

sediment but where the depth was shallow enough to allow for the deployment of kick nets. Benthic invertebrate community data from kick net samples were not used to characterize risk to benthic invertebrates because such samples are methodologically different from the samples collected throughout the rest of the LPRSA and in reference areas. Kick nets and grab samplers collect different types of organisms, resulting in different community compositions. In addition, the other two LOEs for the SQT assessment (i.e., toxicity of LPRSA sediment to benthic invertebrates and sediment chemistry) were not evaluated at these three locations.

Compared to samples from other tidal freshwater LPRSA locations, kick net samples had lower abundances (between 91.3 and 1,870 abundance per m^2) but higher diversity (H′ between 2.1 and 2.5), evenness (J′ between 0.7 and 0.8), and richness (between 20.0 and 27.3), as well as lower HBI (between 5.5 and 6.5) and higher SDI (between 4.7 and 7.0) (Appendix K). Kick nets targeted a different composition of invertebrate species than grab samples. For example, more trichopterans (by abundance) were collected in the three kick net samples than in all other tidal freshwater grab samples combined (Windward 2014).[7]

2.3 COMPARISON TO REFERENCE AREA DATA

Benthic invertebrate community data were evaluated by comparing LPRSA benthic community metrics with those calculated from the reference area datasets described in Section 4.2 of the BERA main text. Reference area data are intended to provide a representation of conditions that would be expected in the LPRSA had the release of site-related hazardous substances not occurred. The use of reference data in this appendix is consistent with USEPA guidance on the use of reference data in ecological risk assessments (ERAs) (USEPA 2002, 2005a).

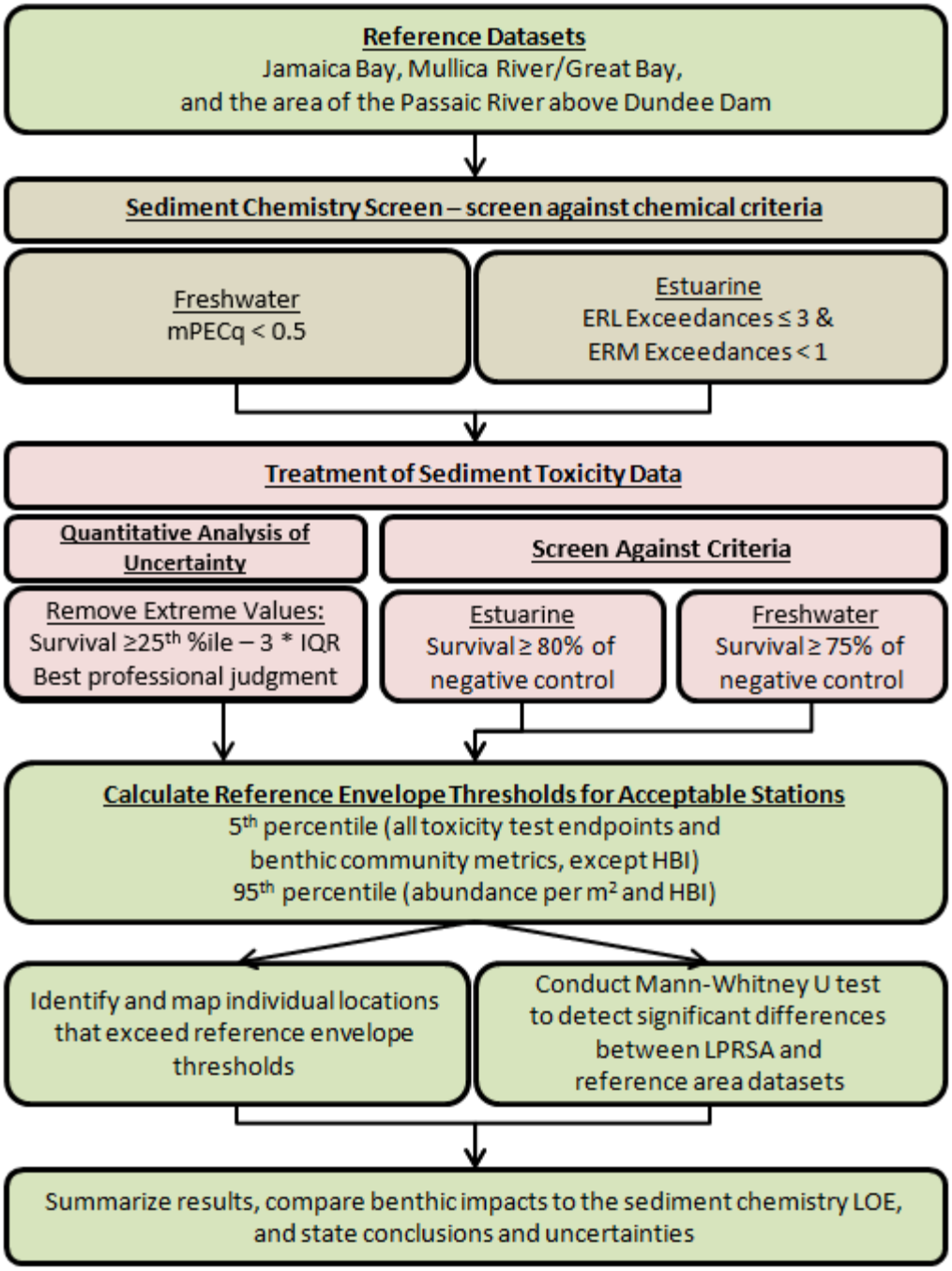
The steps outlined in Figure 2-1 were followed to establish reference datasets and then to compare the LPRSA data with the reference data. Two approaches are described in in Figure 2-1: the approach used in the WOE analysis, and an approach to quantitatively assess uncertainty associated with the approach. The quantitative analysis is presented in the uncertainty sections (Sections 2.3.4, 3.2.4, 4.1.4, and 4.3). Reference area samples were screened using sediment chemistry and toxicity test results. These steps are shown in Appendix B, Tables B3 (estuarine reference area data) and B4 (freshwater reference area data).

Once the reference datasets were fixed, the following analyses were conducted to compare the LPRSA and reference datasets:

[7] Although the higher abundance of trichopterans in kick net samples than in grab samples may have been due in large part to methodological differences between the two sampling methodologies, community compositions are also likely divergent in kick net samples because of significant habitat differences. Specifically, kick net samples were used in highly coarse, shallow areas, whereas other areas tended to have finer sediments capable of penetration by the grab sampler.

- Non-parametric Mann-Whitney U test (one-tailed,[8] alpha [α] = 0.05) to identify significant differences between the LPRSA and reference datasets. A two-tailed version of the test was used to evaluate significant differences between LPRSA and reference abundance data, which can either increase or decrease in response to stress (e.g., decrease in sensitive taxa and/or increase in tolerant, opportunistic taxa). The results of this analysis are presented in Appendix B, Table B6.
- Determine which, if any, locations in the LPRSA exhibited results (either benthic community metrics or toxicity test data) beyond a statistical reference dataset "envelope" (thereby indicating a stressed community) (i.e., less than the 5th percentile reference value [except HBI] and greater than the 95th percentile of abundance or HBI). The calculation of reference envelope thresholds, and the comparison of LPRSA SQT data to those thresholds, are provided in Appendix B, Tables B3 and B4.

[8] One-tailed tests were used to discern significantly lower benthic metric or sediment toxicity test endpoint values in the LPRSA relative to reference area data with two exceptions. HBI, which increases with increasing stress tolerance, was tested using a one-tailed test to discern significantly higher LPRSA HBI values relative to reference area data. Abundance (per m^2) can be elevated in relatively clean habitats or in somewhat stressed habitast. A two-tailed test was used to discern whether LPRSA abundance (per m^2) data were significantly different (higher or lower) than reference area data.



Note: Locations that failed sediment chemistry and/or sediment toxicity screens were removed from the reference area datasets. For the quantitative analysis of uncertainty, locations with extreme toxicity test results were removed from the reference area datasets. These steps are described in more detail in the text. LPRSA locations were defined as freshwater or estuarine if their interstitial salinities (measured during sampling of sediment used to test toxicity) were < 5 ppt or ≥ 5 ppt, respectively.

Figure 2-1. Process for comparing LPRSA benthic datasets with reference datasets

2.3.1 Reference data

Data from four reference locations (i.e., Jamaica Bay, Mullica/Great Bay estuary, freshwater Mullica River, and the area in the Passaic River above Dundee Dam; see Section 4.2 of the BERA main text for a description of the reference areas and Appendix L for the reference data) were compiled to compare with the LPRSA benthic community data (Appendix B, Tables B3-1 and B4-1). Data from these reference locations were also compiled to evaluate LPRSA sediment toxicity data. The reference locations were recommended by USEPA and were agreed upon by CPG, as documented in the Revised RARC Plan (Windward and AECOM Draft). Each of the reference location datasets was used to independently evaluate the LPRSA data. Only those data that passed both the sediment chemistry and sediment toxicity screening criteria (Appendix B, Tables B3-2, B3-3, B4-2, and B4-3) were used to evaluate SQT LOEs. Screening criteria are described below.

Jamaica Bay and the area of the LPR above Dundee Dam represent areas that receive pollution from urban sources but that are not contaminated by the key contaminants associated with the LPRSA. Thus, sediment data from these areas provide a reasonable model for expected future sediment quality conditions in the LPRSA, assuming that sediment risk will be reduced to acceptable levels. Mullica River/Great Bay is a non-urban reference area that represents relatively uncontaminated conditions. Although less relevant to the LPRSA due to its heavily urbanization, the Mullica River/Great Bay reference area provides some understanding of what species could potentially be present in a northeast estuary in the absence of contamination (both associated with Site-specific releases and urban sources). All three reference areas were selected due to their relative proximity to the LPRSA. A comparison of the habitat conditions (i.e., TOC, sediment grain size, and salinity) associated with LPRSA and reference areas is provided in Appendix L (Attachment L2). Jamaica Bay and Mullica River/Great Bay area SQT data were generated using methods similar to those used by CPG to generate SQT data for the LPRSA and the area above Dundee Dam (USEPA REMAP 2009; Windward 2012, 2010, 2009).

Only sediment samples with full SQT data (i.e., sediment chemistry, sediment toxicity, and benthic invertebrate community data) were acceptable for comparisons to LPRSA data. Samples with only benthic invertebrate community data or only sediment chemistry and toxicity data were unacceptable. Moreover, comparable SQT data were required, such that sediment toxicity test results for the same test organism were available in both the LPRSA and reference area datasets. Based on this initial screening criteria, much of the Mullica River freshwater dataset, which only contains freshwater toxicity tests conducted with *Ampelisca abdita*, were removed from consideration.[9] The remainder of the Mullica River freshwater dataset included only benthic invertebrate community data and so was also removed. Thus, all non-urban reference comparisons

[9] Mullica River freshwater data are discussed in the uncertainty section (Section 2.3.4).

to LPRSA data were limited to upper and fluvial estuarine data (i.e., LPRSA locations downstream of RM 13).

Reference area SQT data were screened using sediment chemistry criteria. Sediment chemistry data for estuarine reference locations (Jamaica Bay and Mullica River/Great Bay) were compared to effect range-low (ERL) and effect range-median (ERM) values (Long and Morgan 1990), and sediment chemistry data for freshwater reference locations (i.e., the area above Dundee Dam) were compared to probable effects concentrations (PECs) (MacDonald et al. 2000) (Appendix B, Tables B3-2 and B4-2). Although benthic invertebrate community data were available for many samples from the freshwater portion of the Mullica River reference area, no chemical or toxicological data were available for those same samples; consequently, they were removed from the Mullica River/Great Bay dataset because they could not be screened for acceptability.

Acceptable estuarine reference locations had three or fewer exceedances of ERLs and no exceedances of ERMs across all chemicals for which those sediment guidelines were available. Acceptable freshwater reference locations had mean PEC quotients (mPECqs) (i.e., the average quotient of reference location chemical concentrations and respective PECs) less than 0.5. The method for calculating mPECqs was amended slightly from the method presented by MacDonald et al. (2000) (based on an example calculation provided to CPG by USEPA Region 2): PEC quotients were averaged first within chemical groups (i.e., metals, polycyclic aromatic hydrocarbons [PAHs], organochlorine pesticides, and polychlorinated biphenyls [PCBs]) and then averaged across groups. Averaging PEC quotients in this way had the effect of down-weighting the influence of metal and organochlorine pesticide PEC quotients (n = 8 and n = 6 PECqs, respectively) relative to those for PAHs and PCBs (each based on a single PECq) in the final PECq average.

In addition to meeting the chemical criteria described above, acceptable reference locations were required to meet sediment toxicity criteria. *A. abdita* survival results at estuarine reference locations were required to be ≥ 80% of the respective negative control response. For freshwater reference locations, both *Chironomus dilutus* and *Hyalella azteca* survival results were required to be ≥ 75% of respective negative control responses. These screening steps are shown in Appendix B, Tables B3-3 and B4-3.

Screening the reference data resulted in the removal of several locations from the datasets for Jamaica Bay (n = 59) and the area above Dundee Dam (n = 19) (Appendix B, Tables B3-3 and B4-3). No estuarine SQT locations from Mullica River/Great Bay were removed as a result of the various screening steps described in this section, but all Mullica River freshwater locations (n = 17) were removed as a result of the initial screening step (as described above). The final datasets used to evaluate LPRSA data are discussed in the following sections.

2.3.1.1 Estuarine reference data

LPRSA benthic community data from the upper and fluvial estuarine salinity zones were compared with acceptable data from both Jamaica Bay, which was considered representative of urban habitat conditions, and the Mullica River/Great Bay estuary, which was considered representative of rural habitat conditions. These data are from samples that were collected and analyzed by others (USEPA 2011). Reference area data were used to calculate benthic community metrics (see Appendix L for reference area data and Section 2.1 for methods).

The Jamaica Bay reference dataset consisted of 35 acceptable samples with co-located benthic community, sediment toxicity (*A. abdita* 10-day survival), and sediment chemistry data. Table 2-4 presents a statistical summary of the benthic community metrics from Jamaica Bay. The composition of major taxa from all Jamaica Bay samples, which is dominated by crustaceans and annelid worms, is provided in Figure 2-2.

Table 2-4. Summary of benthic community metrics in the Jamaica Bay estuary reference dataset

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Sample size	35	35	35	35	35
Minimum	410	6.5	0.519	0.232	1.0
Maximum	146,750	48	2.76	0.856	8.0
Mean	20,815	24	1.65	0.594	3.8
Standard deviation	27,728	12	0.543	0.152	1.8
5^{th} percentile	606	7.4	0.774	0.393	1.4
10^{th} percentile	675	8.2	0.906	0.409	1.7
25^{th} percentile	4,269	15	1.40	0.484	2.5
Median	14,050	24	1.66	0.608	4.0
75^{th} percentile	23,850	30	1.99	0.721	4.8
90^{th} percentile	41,960	42	2.33	0.805	6.0
95^{th} percentile	62,809	45	2.43	0.852	7.0

Source: USEPA REMAP (2002b)
REMAP – Regional Environmental Monitoring and Assessment Program
SDI – Swartz's dominance index
USEPA – US Environmental Protection Agency

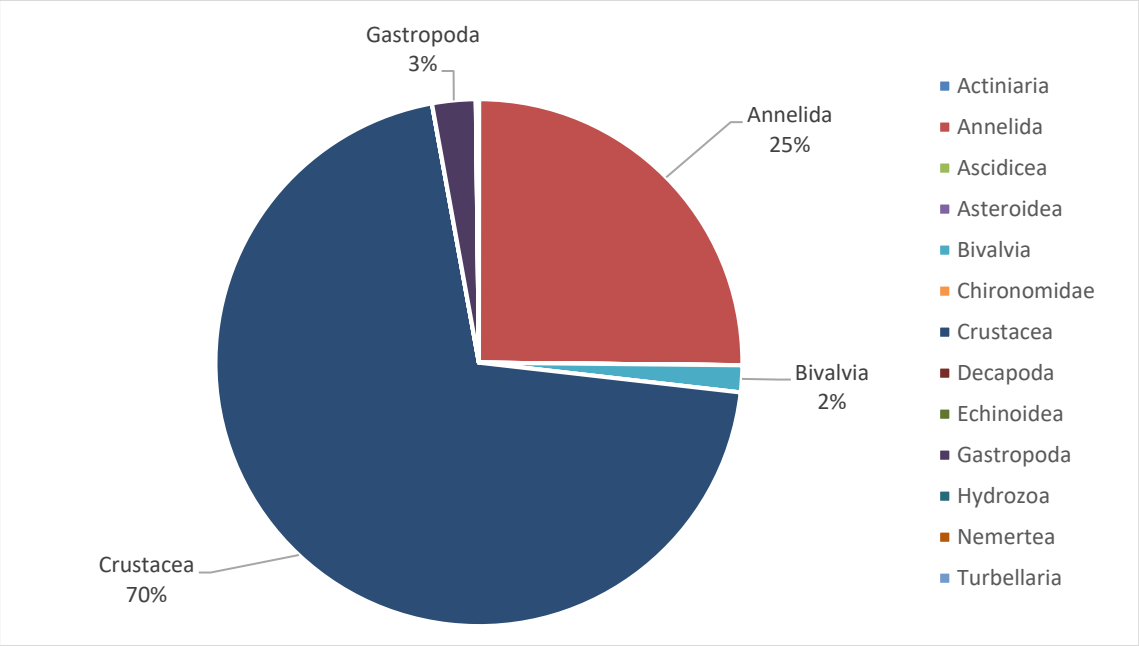


Figure 2-2. Distribution of benthic community relative abundance among major taxa in Jamaica Bay

The Mullica River/Great Bay reference dataset consisted of 12 acceptable samples with co-located benthic community, sediment toxicity, and sediment chemistry data. Table 2-5 presents a statistical summary of the benthic community metrics from Mullica River/Great Bay, and Figure 2-3 shows the community composition among all Mullica River/Great Bay samples based on average major taxa abundance. Like the Jamaica Bay community, the Mullica River/Great Bay community is composed primarily of crustaceans and annelid worms.

Table 2-5. Summary of benthic community metrics in the Mullica River/Great Bay estuary reference dataset

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Sample size	12	12	12	12	12
Minimum	1,470	8	0.92	0.287	2.0
Maximum	51,375	42	3.01	0.884	11
Mean	13,404	25	1.95	0.631	5.1
Standard deviation	17,483	12	0.750	0.202	3.3
5^{th} percentile	1,583	10	0.969	0.332	2.0
10^{th} percentile	1,708	11	1.02	0.372	2.0
25^{th} percentile	2,394	14	1.41	0.432	2.8
Median	5,913	26	1.85	0.705	3.5
75^{th} percentile	12,981	35	2.65	0.777	7.5

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H'	Pielou's J'	SDI (No. of Taxa)
90th percentile	43,698	39	2.91	0.810	9.9
95th percentile	48,474	40	2.96	0.844	11

SDI – Swartz's dominance index

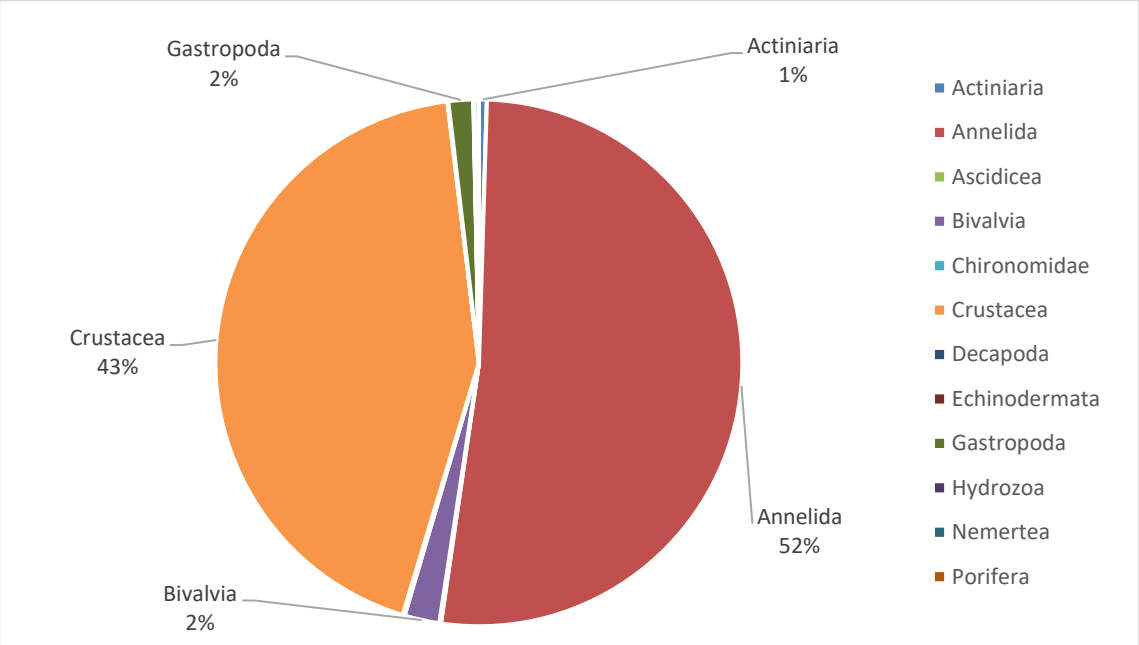


Figure 2-3. Distribution of benthic community relative abundance among major taxa in Mullica River/Great Bay

2.3.1.2 Freshwater reference data

Benthic community data from the LPRSA tidal freshwater zone (RM 13 to RM 17.4) were compared with reference data collected from the Passaic River above Dundee Dam, an area considered representative of sediment quality in the upper section of the LPRSA without the releases of hazardous substances associated with the LPRSA.

As discussed in Section 4.2 of the BERA main text, the reference dataset from the area above Dundee Dam was collected by CPG in fall 2012 and consisted of 24 SQT samples. Sample collection and analysis were conducted following the same methodologies as those used for the collection of samples from the LPRSA. This dataset was screened and reduced according to the approaches described above (and shown in Appenix B, Tables B4-2 and B4-3). After screening, the dataset contained only five samples with co-located sediment chemistry, sediment toxicity test, and benthic invertebrate community metric data. Due to its small size, there is substantial uncertainty associated with this dataset and, by extension, the characterization of reference conditions above Dundee Dam.

Table 2-6 summarizes the benthic community reference data (after screening) for the area above Dundee Dam, and Figures 2-4 and 2-5 show the community composition

among all samples based on average major taxa abundance. The community above Dundee Dam is composed primarily of chironomids and annelid worms.

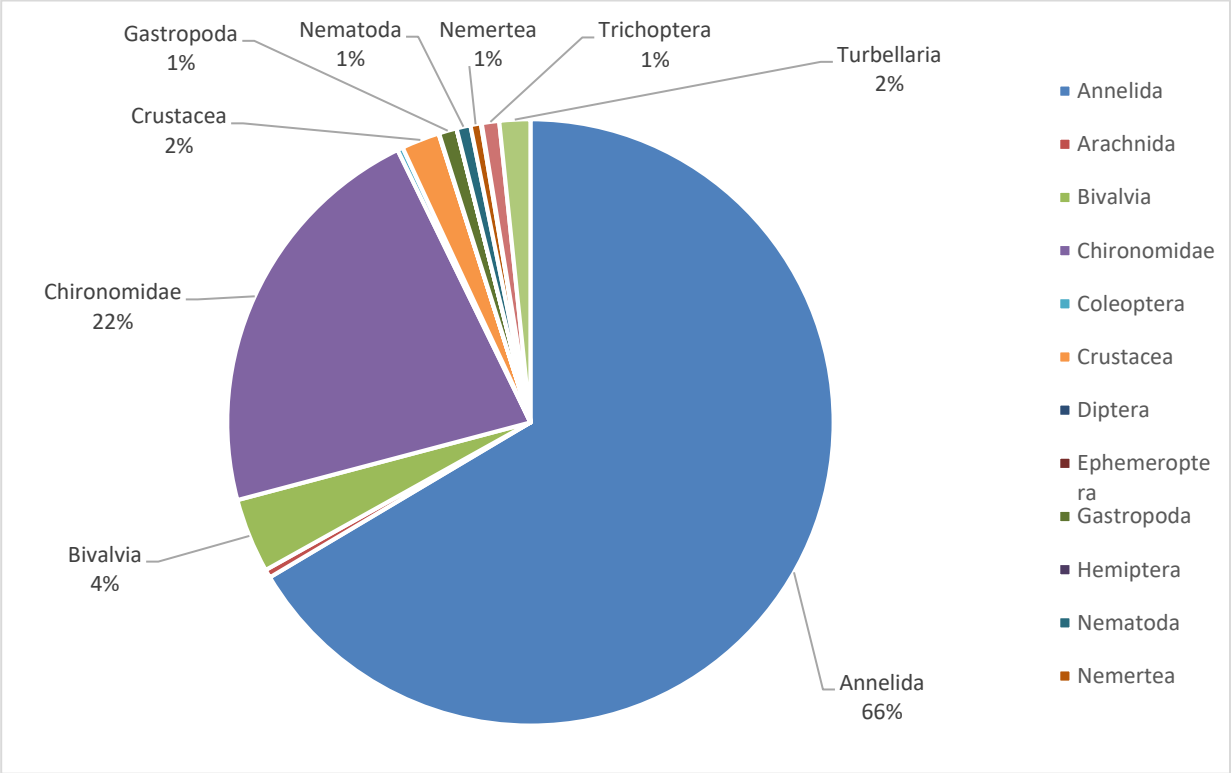


Figure 2-4. Distribution of benthic community relative abundance among major taxa above Dundee Dam

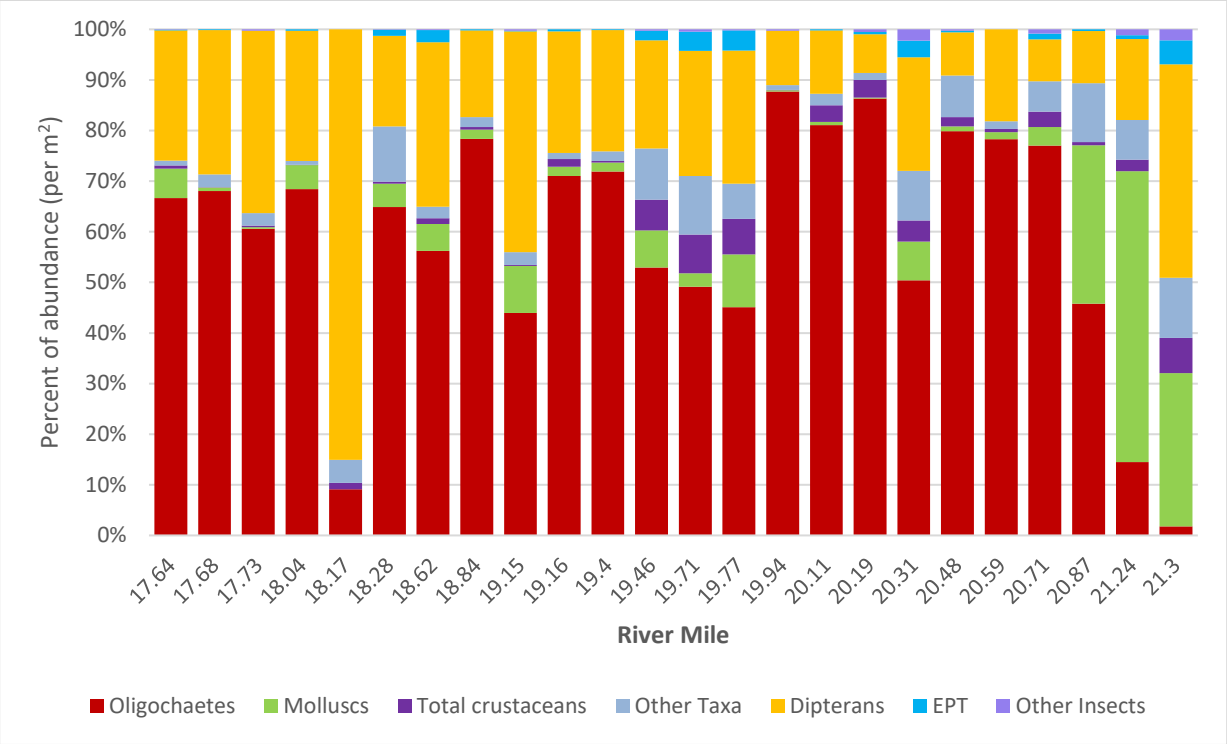


Figure 2-5.Relative abundance of major taxa above Dundee Dam

Table 2-6. Summary of benthic community metrics from the area above Dundee Dam

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI[a]
Sample size	5	5	5	5	5	5
Minimum	11,200	19	1.94	0.617	4.33	8.28
Maximum	19,500	27	2.27	0.727	6.00	9.41
Mean	14,460	22	2.11	0.689	4.93	8.77
Standard deviation	3,453	3	0.133	0.0438	0.833	0.452
5^{th} percentile	11,260	19	1.95	0.630	4.33	8.30
10^{th} percentile	11,320	19	1.97	0.643	4.33	8.33
25^{th} percentile	11,500	19	2.01	0.683	4.33	8.40
Median	14,000	21	2.13	0.697	4.33	8.82
75^{th} percentile	16,100	23	2.19	0.720	5.67	8.94
90^{th} percentile	18,140	25	2.24	0.724	5.87	9.22
95^{th} percentile	18,820	26	2.25	0.726	5.93	9.32

Source: Windward (Draft)

[a] HBI was determined using NYSDEC tolerance values, with the exception of values for arthropods, which were taken from Hilsenhoff (1987).

HBI – Hilsenhoff Biotic Index
SDI – Swartz's dominance index
NYSDEC – New York State Department of Environmental Conservation
Windward – Windward Environmental LLC

2.3.2 Comparison with reference data

The LPRSA benthic community metrics data were compared with reference area data in two ways: by using Mann-Whitney U tests (Appendix B, Table B6) and by identifying LPRSA metric values that were less than the 5th percentile reference value (excepting HBI) or greater than the 95th percentile reference value for abundance and HBI (Appendix B, Tables B3 and B4). Mann-Whitney U tests are used to discern differences between whole datasets, whereas the latter approach discerns location-by-location differences from the majority of reference data. The 5th percentile (and 95th percentile for abundance and HBI) reference values are presented in Table 2-6. Table 2-7 summarizes the results of the Mann-Whitney U tests comparing the benthic invertebrate community data within the LPRSA benthic salinity zones with reference area datasets.

Table 2-7. Results of the statistical analysis comparing the corresponding LPRSA benthic invertebrate community data with reference data

Benthic Salinity Zone	Area of Interest	Is There a Significant Difference Between LPRSA Reference Area Data? (Mann-Whitney U test, alpha = 0.05)					
		Abundance (per m^2)[a]	Taxa Richness	Shannon-Weiner H'	Pielou's J'	SDI	HBI[b]
Upper estuary	Jamaica Bay (urban)	**yes; 8.74E-07**	**yes; 7.65E-08**	**yes; 2.07E-02**	no; 9.86E-01	**yes; 3.16E-02**	na[c]
Fluvial estuary	Jamaica Bay (urban)	**yes; 2.20E-03**	**yes; 2.45E-09**	**yes; 8.60E-07**	**yes; 2.69E-02**	**yes; 4.86E-07**	na[c]
Upper estuary	Mullica River and Great Bay (non-urban)	**yes; p = 1.28E-05**	**yes; p = 1.06E-05**	**yes; p = 2.30E-02**	no; p = 6.69E-01	no; p = 6.27E-02	na[c]
Fluvial estuary	Mullica River and Great Bay (non-urban)	no; p = 2.35E-01	**yes; p = 7.97E-06**	**yes; p = 1.40E-04**	**yes; p = 4.26E-02**	**yes; p = 7.95E-05**	na[c]
Tidal freshwater	Above Dundee Dam (urban)	no; p = 8.64E-02	**yes; p = 2.57E-03**	**yes; p = 9.85E-04**	**yes; p = 4.06E-03**	**yes; p = 1.98E-03**	no; p = 1.40E-01

Note: **Bold** text indicates significant result ($p < 0.05$); significance/p-values are based on lower-tailed Mann Whitney U tests, unless otherwise noted.

[a] Abundance tested using two-tailed Mann-Whitney U test.

[b] HBI tested using upper-tailed Mann-Whitney U test.

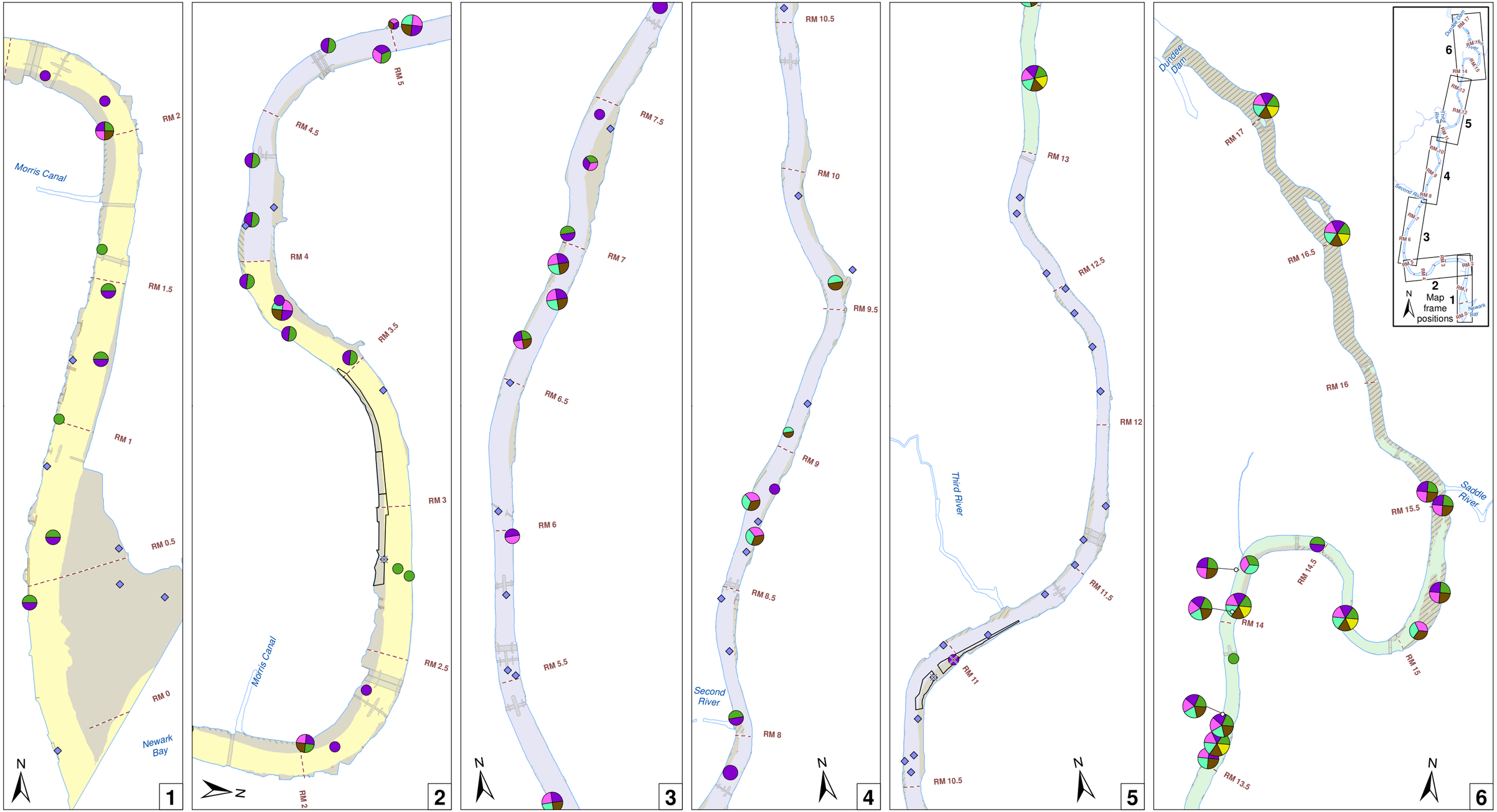
[c] HBI is specific to freshwater, so comparison to Jamaica Bay or Mullica River/Great Bay data was not appropriate

HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River Study Area
na – not applicable
SDI – Swartz's dominance index

As shown in Table 2-7, results from the Mann-Whitney U test demonstrate statistically significant differences in benthic community metrics between the LPRSA and reference areas. Taxa richness and diversity (H') were consistently lower in the LPRSA than in reference areas, and dominance (SDI) was similarly depressed in the fluvial

estuarine and tidal freshwater zones relative to reference conditions. Evenness (J') was significantly different between the tidal freshwater LPRSA and the area above Dundee Dam, as well as between the fluvial estuarine LPRSA and each estuarine reference area. Results were similar between the urban and non-urban comparisons in the estuarine zones of the LPRSA (Table 2-7). These results indicate that, in general, the benthic invertebrate community in the LPRSA is impaired relative to communities found in reference areas.

Figure 2-6 shows the results of comparisons (by sample location) of LPRSA benthic community metrics with reference envelope thresholds (Appendix B, Tables B3 and B4). The comparison shown in Figure 2-6 is based on Jamaica Bay data compared with data from locations in the upper estuarine and fluvial estuarine benthic salinity zones of the LPRSA, and data from above Dundee Dam compared with data from locations in the tidal freshwater zone of the LPRSA, respectively. The number and frequency of LPRSA locations with data different from the range of reference area values are summarized in Table 2-8. Figure 2-7 presents similar results based on the comparison of LPRSA data to non-urban Mullica River and Great Bay data.



Metric outside reference envelope[a]

- Abundance (per m²)
- Taxa Richness
- Shannon-Wiener H' (log e)
- Pielou's J'
- Swartz's Dominance
- Hilsenhoff Biotic Index

- All metrics within reference envelope
- Remediated location[b]
- Bridge
- Abutment
- Dock
- Dredge zone

- River mile
- Mudflat
- Gravel with fines
- Upper Estuarine
- Fluvial Estuarine
- Tidal Freshwater

0 0.25 0.5 Miles

0 0.25 0.5 0.75 Kilometers

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

[a]Hilsenhoff Biotic Index value above the 95th percentile of the reference data; Abundance value above the 95th percentile or below the 5th percentile; all other metrics below the 5th percentile. Note: Locations downstream of RM 13 were compared to urban reference data from Jamaica Bay; locations upstream of RM 13 were compared to urban reference data from the area of the Passaic River above Dundee Dam.

[b]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

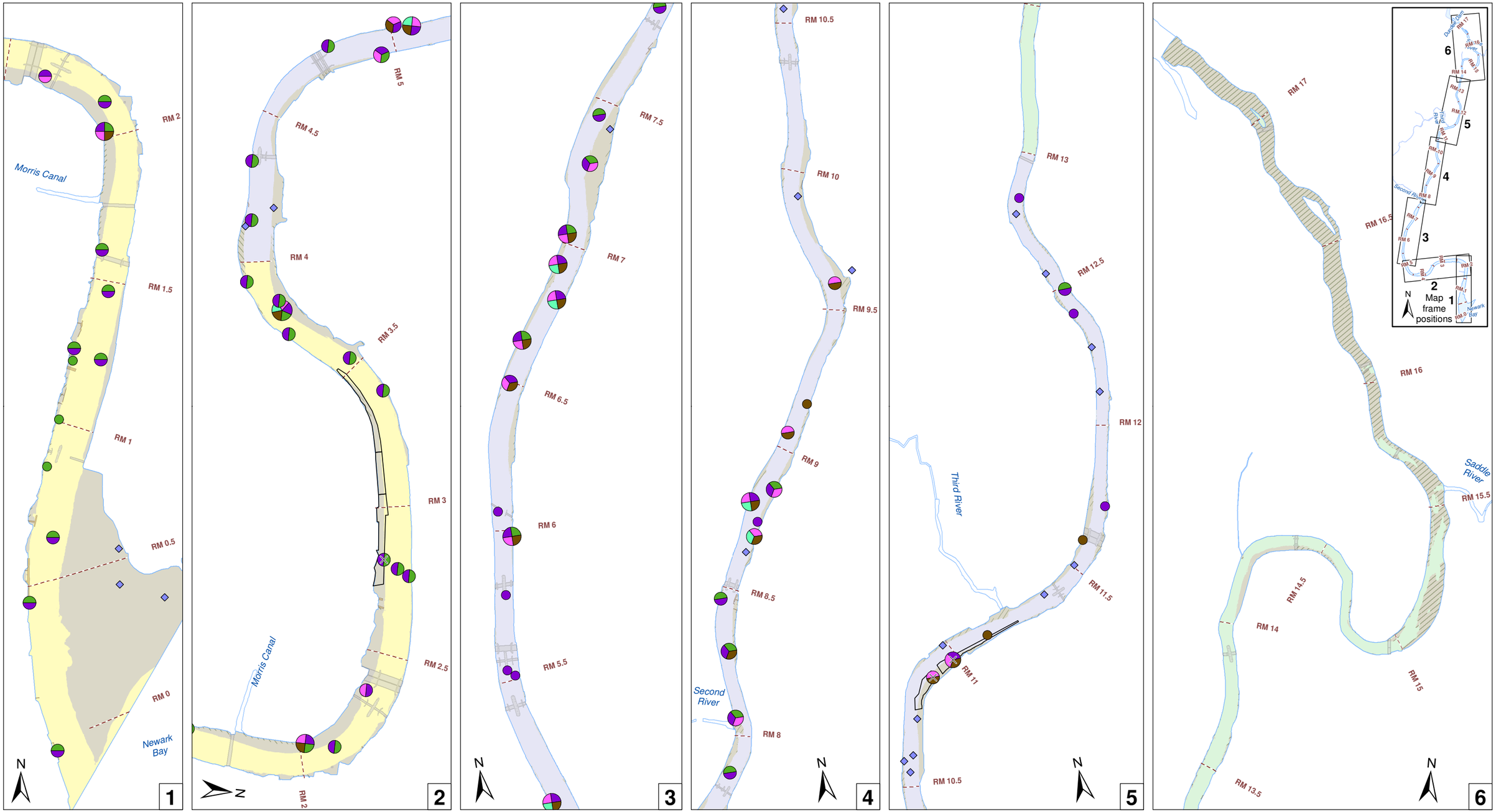
Figure 2-6. Benthic community metrics in the LPRSA compared with reference data representing urban habitats

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Windward environmental LLC

Prepared by mikey 4/23/2019 ; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6369_Benthic metrics_Urban_LSM_20160721.mxd



Metric outside reference envelope[a]

- Abundance (per m²)
- Taxa Richness
- Shannon-Weiner H' (log e)
- Pielou's J
- Swartz's Dominance

- All metrics within reference envelope
- Remediated location[b]
- Bridge
- Abutment
- Dock
- Dredge zone

- River mile
- Mudflat
- Gravel with fines
- Estuarine LPRSA
- Transitional LPRSA
- Freshwater LPRSA

0 0.25 0.5 Miles

0 0.25 0.5 0.75 Kilometers

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

[a]Abundance value above the 95th percentile or below the 5th percentile; all other metrics below the 5th percentile of the reference data. Note: Locations were compared to non-urban reference data from Mullica River and Great Bay.

[b]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

Windward environmental LLC

Figure 2-7. Benthic community metrics in the LPRSA compared with reference data representing non-urban habitats

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6441_Benthic metrics_nonUrban_LSM_20160721.mxd

Table 2-8. Summary of comparison of benthic metrics with reference envelope

Benthic Salinity Zone	Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI[a]
Comparison with Jamaica Bay (urban)							
Upper estuary (RM 0 to RM 4)	sample size	25	25	25	25	25	na[b]
	no. of locations outside reference envelope	12	12	2	1	2	na[b]
	percent of locations outside reference envelope	48%	48%	8%	4%	8%	na[b]
Fluvial estuary (RM 4 to RM 13)[c]	sample size	54	54	54	54	54	na[b]
	no. of locations outside reference envelope	8	19	10	7	9	na[b]
	percent of locations outside reference envelope	15%	35%	19%	13%	17%	na[b]
Comparison with Mullica River and Great Bay (non-urban)							
Upper estuary (RM 0 to RM 4)	sample size	25	25	25	25	25	na[b]
	no. of locations outside reference envelope	21	19	3	1	2	na[b]
	percent of locations outside reference envelope	84%	76%	12%	4%	8%	na[b]
Fluvial estuary (RM 4 to RM 13)[c]	sample size	54	54	54	54	54	na[b]
	no. of locations outside reference envelope	15	30	18	5	18	na[b]
	percent of locations outside reference envelope	28%	56%	33%	9%	33%	na[b]
Comparison with the area above Dundee Dam (urban)							
Tidal freshwater (RM 13 to RM 17.4)	sample size	18	18	18	18	18	18
	no. of locations outside reference envelope	16	15	16	12	15	6
	percent of locations outside reference envelope	89%	83%	89%	67%	83%	33%

[a] HBI was determined using NYSDEC tolerance values. HBI was the only metric compared to the maximum; all others were compared to the minimum.

[b] HBI applies to only tidal freshwater locations.

[c] Methods of sample collection differed between these two datasets (e.g., sieve sizes used were 1 mm or 0.5 mm); within the fluvial estuarine zone of the LPRSA, only samples upstream of RM 8.5 differed in this way.

HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River Study Area
na – not applicable
NYSDEC – New York State Department of Environmental Conservation
RM – river mile
SDI – Swartz's dominance index

2.3.3 Summary of results

Using the reference envelope threshold values to establish differences between LPRSA benthic metrics values and reference conditions, several sampling locations

were outside the reference envelope (Table 2-8). For example, 48% of the 25 locations in the upper estuarine zone of the LPRSA were outside the urban reference envelope for both richness and abundance per m^2. By comparison, evenness, diversity, and dominance were more often similar to urban reference conditions at these locations (i.e., in ≤ 8% of 25 samples). The results of multiple Mann-Whitney U tests indicate that metrics in the upper estuarine zone are generally different than metrics in Jamaica Bay (the only exception being evenness).

In the fluvial estuarine salinity zone, richness was again the most sensitive metric, followed by abundance and diversity (Table 2-8). Evenness and dominance reference condition thresholds were exceeded by a smaller margin in the fluvial estuarine zone than in the upper estuarine zone. However, based on Mann-Whitney U tests, metrics tended to be less in the LPRSA upper fluvial estuarine zone than in the urban reference area (Table 2-7).

Benthic community metrics from the tidal freshwater zone of the LPRSA were generally dissimilar to metrics in the area above Dundee Dam (Tables 2-7 and 2-8). For example, abundance, richness, diversity, and dominance all exceeded the reference envelope at more than 80% of tidal freshwater locations, and all metrics except abudance were generally lower than reference conditions based on Mann-Whitney U tests. Only HBI was generally similar (on a sample-by-sample basis), with only 33% (n = 6) of tidal freshwater SQT locations falling outside the urban freshwater reference envelope.

2.3.4 Uncertainties in comparison to reference area data

This section identifies the uncertainties inherent in the benthic invertebrate community LOE analyses provided above. Uncertainties are as follows:

- It is unclear whether the screened reference area datasets accurately reflect the reference condition. Specifically, screening reference area data using sediment chemistry or sediment toxicity criteria imposes a potentially unreasonable constraint on data acceptability. The resulting datasets may not capture the full range of possible benthic community metric results that should be expected under urbanized conditions (but for the LPRSA-specific release of hazardous materials). This is particularly true of the dataset from above Dundee Dam, which, after screening, contained only five acceptable SQT samples.
- It is unknown whether temporal factors influence the interpretation of reference area datasets. These factors could include annual or seasonal changes in invertebrate communities or community responses to significant events such as storms, discharge events, droughts, or seasonal disturbances. Within reference area datasets, temporal changes could result in invertebrate data variability.
- It is unclear whether comparing LPRSA data to data from a non-urban reference area (Mullica River and Great Bay) is relevant for characterizing risks

in the LPRSA. Comparison to non-urban conditions fails to incorporate potential stressors that are generally observed in urban settings and are expected to influence the LPRSA benthic invertebrate community. Examples of these stressors include altered hydrology due to channelization and flood controls and increased organic and inorganic inputs from CSOs, SWOs, road waste, and permitted industrial discharges.

- Abundance (per m^2) and taxa richness (as well as the metrics that are based on these measurements) are influenced by sample volume and size; the variability in sampling methods upstream and downstream of RM 8.5 (as well as those in reference areas) may have influenced these measurements and subsequent analyses. The difference in sieve sizes used to collect benthic invertebrate community samples downstream of RM 8.5 (1 mm), upstream of RM 8.5 (0.5 mm) (including the area above Dundee Dam), and in Jamaica Bay (0.5 mm) could have influenced sample abundance and taxa richness (Gage et al. 2002). Therefore, comparisons made between data from the tidal freshwater LPRSA and data from a portion of the fluvial estuarine zone of the LPRSA between RM 8.5 and RM 13 (which were compared to Jamaica Bay) may be influenced by differences in mesh size. This uncertainty does not apply to comparisons made between data from locations above Dundee Dam and within the tidal freshwater zone of the LPRSA.
- Both abundance and taxa richness may increase with increasing environmental stress under certain circumstances (Pearson and Rosenberg 1978). Based on the frequency and intensity of disturbance in the LPRSA, late mature communities are not likely to exist in the LPRSA, so disturbance is unlikely to result in increased taxa richness (consistent with the Pearson-Rosenberg model). Abundance may increase in disturbed habitats of the LPRSA, driven by relatively few disturbance-tolerant taxa. The evaluation of risk relative to reference conditions from urban habitats based on abundance (per m^2) was somewhat uncertain for that reason. However, the use of both 5^{th} and 95^{th} percentiles as reference thresholds, as well as the use of a two-tailed Mann-Whitney U test for that variable (Table 2-7), is intended to account for this uncertainty.
- Although the area downstream of RM 4 was assumed to have a spatially consistent daily estuarine salinity (i.e., ≥ 5 parts per thousand [ppt]), this area is still prone to seasonal shifts in salinity (i.e., estuarine to complete freshwater [< 0.5 ppt]) (Moffatt & Nichol 2013). Therefore, disturbances observed in this area during seasonal low flows (such as the flows preceding the 2009 SQT sampling event) (USGS 2012) may be related to the seasonal influx of estuarine waters into what was previously freshwater habitat. In other words, substantial shifts in the benthic invertebrate community from predominately freshwater species to estuarine species may occur from RM 0 to RM 4 over the course of a

year (Windward 2014), and this seasonal shift may be responsible, in part, for the disturbed communities observed in the upper estuarine zone of the LPRSA. Intermittent physical disturbance is also a possibility; significant deposition and erosion in the upper estuarine zone were observed during a sediment profile imaging survey in June 2005 (Germano & Associates 2005).

- It is possible that significant differences between freshwater LPRSA benthic community metrics and metrics above Dundee Dam are related to differences in sediment grain size between the tidal freshwater LPRSA (very coarse in some places) and the area above Dundee Dam (finer). Alternatively, differences may be related to the influence of Dundee Dam (i.e., unnatural hydrology).
- Though the Mullica River freshwater dataset was insufficient to characterize a non-urban freshwater reference condition, the ranges of community metric data from the three acceptable Mullica River freshwater samples are presented in Table 2-9, and the composition of major taxa (based on average abundance across all samples) is shown in Figure 2-8.[10] Ranges of Mullica River freshwater metric data generally overlap with ranges from above Dundee Dam (and the LPRSA). Slightly lower minimum and maximum values in the Mullica River freshwater dataset suggest that the comparison of LPRSA data to data from the area above Dundee Dam is conservative. HBI is an exception, in that the much lower range in Mullica River freshwater data indicates a far less tolerant community at that non-urban site. This is corroborated by Figure 2-8, in that sensitive taxa (e.g., Crustacea, Ephemeroptera, and Trichoptera) compose a larger portion of the freshwater Mullica River than the urban reference area.
- Benthic community metrics can be calculated in more than one way, depending on how replicate community samples and metrics are averaged. For the analysis herein, benthic community metrics were calculated for each sample replicate, and the replicate metric values were averaged to obtain a single metric value (and variance) for a sampling location. Alternatively, replicate benthic community data can be averaged together first, followed by the calculation of a single benthic community metric value (without variance) for the entire location. This alternative approach can result in very different metric estimates, generally trending toward greater abundance, richness, diversity, and evenness, and lower dominance and pollution tolerance (Table 2-10). For the LPRSA data, this trend is most pronounced for richness, diversity, and tolerance and least pronounced for abundance, evenness, and dominance. This difference indicates that the analysis of LPRSA data relative to reference area datasets is biased toward predicting greater impacts (i.e., conservative).

[10] One of four Mullica River freshwater locations (NJ06-0046-A) was screened out due to low sediment toxicity; *A. abdita* survival was 41.8% of negative control, which is below the screening criterion of 75% of control.

Table 2-9. Ranges of benthic community metrics from the LPRSA, above Dundee Dam, and the freshwater Mullica River

Area of Interest	Abundance (per m²)	Taxa Richness	Shannon-Wiener H′	Pielou's J′	SDI	HBI
Freshwater LPRSA (RM 14 to RM 17.4)	1,990–30,600	8.0–23	1.32–2.0	0.507–0.730	2.0–5.3	8.33–9.73
Above Dundee Dam[a]	11,200–19,500	19–27	1.94–2.27	0.617–0.727	4.3–6.0	8.28–9.41
Mullica River freshwater[b]	3,900–5,825	14–19	1.53–2.05	0.581–0.713	3.0–5.0	6.79–8.77

[a] Ranges based on screened reference dataset (Table 2-6).

[b] One of four Mullica River freshwater locations (NJ06-0046-A) was screened out due to low sediment toxicity; *A. abdita* survival at that location was 41.8% of negative control, which was below the screening criterion of 75% of control.

HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River study Area
RM – river mile
SDI – Swartz's dominance index

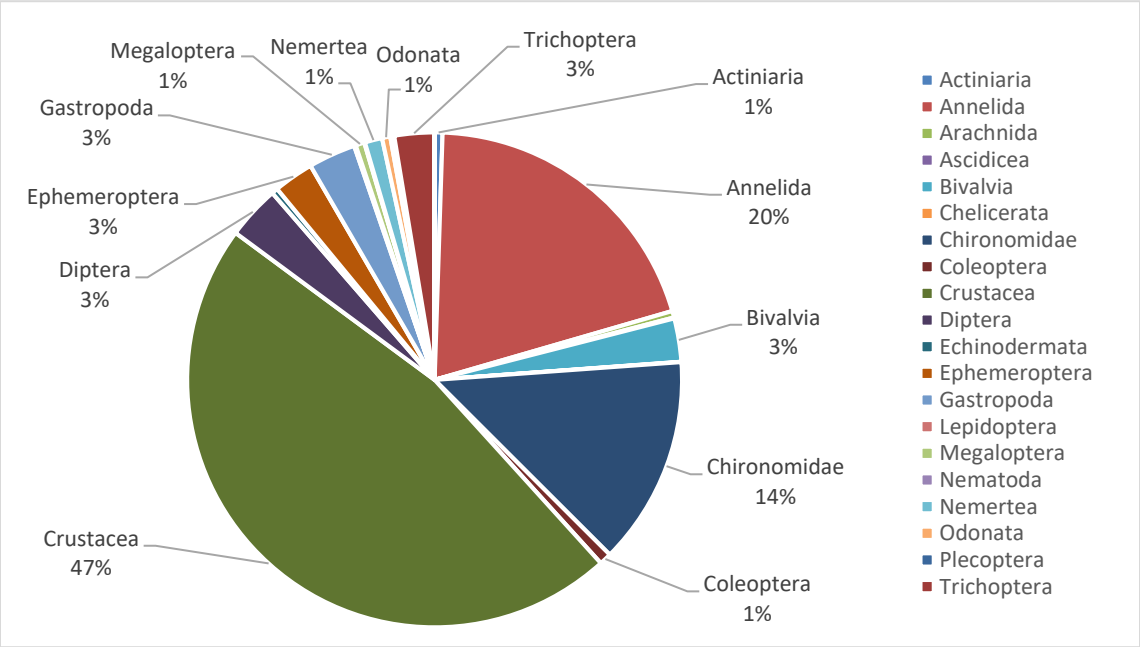


Figure 2-8.Relative abundance of major taxa in the freshwater Mullica River

Table 2-10. Comparison of benthic community metrics calculated using two different averaging methods

Location ID	Abundance (per m²)		Richness		Shannon-Wiener H'		Pielou's J'		SDI		HBI[a]	
	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value
LPRT01A	1,090	1,467	8	14	1.06	1.72	0.507	0.651	2.0	3	--	--
LPRT01B	614	288	8	10	1.34	1.92	0.662	0.834	2.7	5	--	--
LPRT01C	2,740	2,853	16	21	2.01	2.22	0.727	0.729	4.7	5	--	--

Location ID	Abundance (per m^2)		Richness		Shannon-Wiener H'		Pielou's J'		SDI		HBI[a]	
	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value
LPRT01D	2,050	2,132	12	18	1.63	1.83	0.650	0.633	3.0	3	--	--
LPRT01E	2,750	4,418	13	17	1.71	1.81	0.669	0.638	3.1	3	--	--
LPRT01F	357	448	7	13	1.06	1.70	0.537	0.663	2.0	5	--	--
LPRT01G	973	1,078	11	17	1.42	1.79	0.600	0.630	2.3	3	--	--
LPRT02A	709	1,003	9	23	1.57	2.54	0.718	0.811	3.1	7	--	--
LPRT02B	127	213	7	12	1.68	2.32	0.887	0.932	4.0	7	--	--
LPRT02C	630	710	8	10	1.67	1.94	0.823	0.843	3.7	4	--	--
LPRT02D	1,290	933	13	14	2.07	2.32	0.816	0.877	5.0	7	--	--
LPRT02E	155	215	6	9	1.42	1.83	0.830	0.832	3.5	4	--	--
LPRT02F	240	398	9	19	1.73	2.42	0.807	0.821	4.0	10	--	--
LPRT03A	320	413	3	6	0.427	1.07	0.430	0.597	1.3	2	--	--
LPRT03B	773	945	6	9	1.45	1.77	0.803	0.808	3.0	4	--	--
LPRT03C	894	2,767	7	11	1.21	0.583	0.610	0.243	2.6	1	--	--
LPRT03D	1,370	603	11	15	1.69	2.32	0.728	0.858	3.9	6	--	--
LPRT03E	593	850	8	13	1.17	1.73	0.690	0.675	2.0	3	--	--
LPRT03F	693	888	9	13	1.49	2.07	0.697	0.808	3.0	5	--	--
LPRT04A	3,420	1,240	9	12	1.31	1.83	0.599	0.738	2.6	4	--	--
LPRT04B	523	633	6	10	1.41	1.84	0.790	0.801	2.7	4	--	--
LPRT04C	721	222	8	10	1.52	2.17	0.778	0.942	3.4	6	--	--
LPRT04D	1,220	1,343	5	9	0.507	0.841	0.297	0.383	1.0	1	--	--
LPRT04E	1,170	1,193	6	8	1.07	1.31	0.607	0.630	2.3	3	--	--
LPRT04F	333	380	5	8	1.09	1.58	0.757	0.760	2.3	3	--	--
LPRT05A	2,280	4,720	11	17	1.54	1.74	0.640	0.615	3.0	3	--	--
LPRT05B	852	708	9	14	1.57	1.75	0.746	0.664	3.6	3	--	--
LPRT05C	2,940	3,037	10	14	1.33	1.62	0.570	0.615	2.3	3	--	--
LPRT05D	227	297	6	11	1.08	1.86	0.603	0.775	2.7	5	--	--
LPRT05E	200	230	4	6	1.08	1.41	0.850	0.785	2.0	3	--	--
LPRT05F	147	223	3	7	0.737	1.33	0.740	0.683	2.0	2	--	--
LPRT06A	6,530	5,200	7	8	0.843	0.788	0.463	0.379	1.6	1	--	--
LPRT06B	3,380	3,472	4	9	0.147	0.271	0.100	0.123	1.0	1	--	--
LPRT06C	4,660	3,248	10	12	1.20	1.25	0.528	0.503	2.0	2	--	--
LPRT06D	7,030	11,665	8	14	1.13	1.39	0.533	0.528	2.0	3	--	--
LPRT06E	2,490	3,562	7	10	1.21	1.48	0.603	0.641	2.3	3	--	--
LPRT06F	927	933	2	3	0.547	0.480	0.727	0.437	1.7	1	--	--

Location ID	Abundance (per m^2)		Richness		Shannon-Wiener H'		Pielou's J'		SDI		HBI[a]	
	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value
LPRT07A	5,110	4,540	9	13	0.942	1.20	0.436	0.468	1.8	2	--	--
LPRT07B	3,200	3,548	8	12	0.933	1.26	0.467	0.506	1.7	2	--	--
LPRT07C	2,680	670	7	10	0.893	0.980	0.469	0.426	1.4	1	--	--
LPRT07D	1,630	1,758	5	10	0.440	0.734	0.260	0.319	1.0	1	--	--
LPRT07E	3,740	3,773	7	9	0.383	0.438	0.217	0.199	1.0	1	--	--
LPRT08A	253	405	4	9	0.863	1.18	0.657	0.539	1.7	2	--	--
LPRT08B	340	352	3	4	0.773	0.910	0.703	0.656	2.0	2	--	--
LPRT08C	4,030	3,503	11	17	1.23	1.37	0.534	0.484	2.2	2	--	--
LPRT08D	843	882	6	8	1.10	1.20	0.677	0.577	2.0	2	--	--
LPRT08E	4,070	697	8	9	1.04	1.50	0.523	0.682	1.8	3	--	--
LPRT09A	1,300	1,387	9	15	1.14	1.50	0.510	0.554	2.0	2	--	--
LPRT09B	533	1,215	4	9	0.877	0.989	0.717	0.450	2.0	2	--	--
LPRT09C	9,890	11,514	9	15	1.07	1.18	0.550	0.435	2.0	2	--	--
LPRT09D	20,300	8,443	11	15	0.947	0.844	0.399	0.312	1.6	1	--	--
LPRT09E	6,180	6,443	9	15	1.04	1.21	0.460	0.448	2.0	2	--	--
LPRT09F	5,380	5,576	9	13	0.647	0.811	0.300	0.316	1.0	1	--	--
LPRT09G	1,090	1,780	5	10	0.967	1.47	0.627	0.637	2.0	2	--	--
LPRT09H	1,330	2,827	7	13	0.980	1.28	0.503	0.500	1.7	2	--	--
LPRT10A	11,100	11,483	10	16	0.810	0.907	0.357	0.327	1.3	1	--	--
LPRT10B	7,210	7,669	15	21	1.22	1.47	0.453	0.482	1.7	2	--	--
LPRT10C	24,300	10,573	13	22	0.968	1.10	0.378	0.354	1.4	1	--	--
LPRT10D	5,030	6,127	14	24	1.49	2.10	0.593	0.660	3.0	5	--	--
LPRT10E	37,900	14,110	12	16	1.19	1.45	0.484	0.522	2.3	2	--	--
LPRT11A	4,280	4,643	14	21	1.48	1.79	0.570	0.586	2.3	3	--	--
LPRT11B	4,610	5,163	13	20	1.65	1.95	0.647	0.652	3.0	4	--	--
LPRT11C	6,100	6,926	14	23	1.28	1.71	0.497	0.546	2.3	4	--	--
LPRT11D	29,600	17,526	16	26	1.38	1.80	0.500	0.552	2.6	3	--	--
LPRT11E	21,200	6,469	11	19	1.08	1.69	0.462	0.574	2.1	3	--	--
LPRT11F	6,050	6,258	16	22	1.62	1.81	0.587	0.586	3.3	4	--	--
LPRT11G	3,170	3,513	7	12	0.840	1.35	0.440	0.542	1.7	2	--	--
LPRT12A	7,270	4,127	16	16	1.38	1.42	0.510	0.513	2.7	2	--	--
LPRT12B	8,160	9,721	13	21	1.34	1.77	0.520	0.583	2.7	3	--	--
LPRT12C	9,850	11,461	15	32	1.26	1.79	0.463	0.518	2.3	3	--	--
LPRT12D	5,870	6,586	11	16	1.07	1.59	0.443	0.572	1.7	3	--	--

Location ID	Abundance (per m^2)		Richness		Shannon-Wiener H'		Pielou's J'		SDI		HBI[a]	
	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value	BERA Value	Alt. Value
LPRT12E	20,900	5,183	10	11	1.23	1.56	0.570	0.652	2.4	3	--	--
LPRT13A	2,600	2,933	11	22	1.28	1.73	0.540	0.561	2.0	3	--	--
LPRT13B	5,790	6,648	12	20	1.33	1.71	0.547	0.571	2.0	3	--	--
LPRT13C	4,630	4,710	9	12	1.07	1.15	0.490	0.463	2.0	2	--	--
LPRT13D	680	980	8	14	1.43	2.31	0.727	0.877	3.3	6	--	--
LPRT13E	23,200	13,490	13	18	1.40	1.59	0.550	0.551	2.9	3	--	--
LPRT13F	12,400	2,917	13	14	1.42	1.68	0.554	0.637	2.7	3	--	--
LPRT13G	2,690	3,154	8	15	1.15	1.59	0.570	0.587	2.0	3	--	--
LPRT14A	27,600	6,739	14	21	1.41	1.79	0.543	0.588	2.6	3	9.56	6.35
LPRT14B	11,500	12,502	14	21	1.63	1.92	0.620	0.631	3.7	4	9.36	5.56
LPRT14C	9,120	11,191	18	23	1.36	1.84	0.468	0.588	2.3	4	9.39	8.22
LPRT14D	7,980	8,753	17	25	1.73	2.05	0.610	0.635	3.7	5	8.99	7.01
LPRT14E	4,520	4,747	15	20	1.57	1.81	0.587	0.606	2.7	3	8.54	6.84
LPRT14F	8,450	9,125	19	27	1.96	2.15	0.677	0.653	5.0	6	8.67	6.79
LPRT15A	30,600	35,597	19	29	1.61	1.99	0.550	0.590	2.7	4	8.99	5.94
LPRT15B	21,800	5,446	14	30	1.34	2.02	0.562	0.595	2.3	3	9.64	4.69
LPRT15C	2,060	3,460	12	24	1.55	2.17	0.637	0.684	3.3	5	8.53	4.64
LPRT15D	15,700	9,425	17	38	1.69	2.41	0.600	0.663	3.2	8	9.32	5.56
LPRT15E	7,020	4,267	18	20	1.85	2.26	0.640	0.755	4.0	8	9.10	6.63
LPRT15F	4,180	4,463	15	24	1.61	2.04	0.593	0.643	3.3	5	8.76	5.56
LPRT16A	12,800	13,917	23	38	1.58	1.95	0.507	0.537	2.7	4	9.19	7.13
LPRT16C	10,400	7,437	21	24	2.07	2.12	0.710	0.666	5.3	5	8.56	6.68
LPRT16D	8,680	12,627	17	29	1.80	2.37	0.633	0.705	4.0	6	8.06	4.32
LPRT16E	1,990	2,267	8	14	1.32	1.74	0.640	0.661	2.7	3	8.91	3.01
LPRT17A	2,770	3,173	11	18	1.41	1.81	0.583	0.625	2.0	3	8.96	6.93
LPRT17D	3,640	9,358	10	26	0.901	1.63	0.433	0.500	1.7	3	9.45	7.15

Note: Alternative values are based on the alternative approach to calculating benthic community metrics: averaging the abundance of species across location replicates prior to calculating a single metric value. This method is distinguished herein from the method used in the BERA: calculating a metric value for each replicate prior to averaging the metric values across replicates.

[a] HBI is calculated only for samples in the tidal freshwater zone.

BERA – baseline ecological risk assessment
HBI – Hilsenhoff Biotic Index
ID – identification
SDI – Swartz's dominance index

2.3.4.1 Quantitative analysis of uncertainty

A quantitative analysis was conducted to address uncertainties associated with the benthic invertebrate community LOE. Methods and results of the analysis are detailed in this section.

Methods

Reference sediment data were screened differently than as described in Figure 2-1. An analysis of extreme low-toxicity values based on the interquartile range (IQR) was conducted to determine if there were reference area sediment toxicity data that were substantially less than the other reference area data (Appendix B, Tables B3-3 and B4-3). Reference locations were identified as outliers if a toxicity test survival endpoint value was less than three times the IQR below the 25th percentile of the data. When deemed necessary, best professional judgment was used to identify samples with sediment toxicity above the IQR-based threshold, but that were clearly impacted in some way aside from sediment chemical contamination. Although the toxic effects associated with urban sediments may be useful for establishing relevant reference envelope thresholds, allowing for extremely low values (e.g., survival < 25%) is inconsistent with the purpose of the reference condition approach. The extreme value removal step is intended to eliminate reference stations that are uncharacteristic of the reference area dataset (objectively defined) without imposing a predefined toxicity screening criterion (i.e., 75 or 80% survival relative to control). The IQR-based toxicity data screening procedure allows for greater variability in possible reference area data but does not allow for extreme low toxicity. Toxicity in sediments from reference areas may occur for several reasons (e.g., elevated total organic carbon [TOC] or excessive fine-grained sediments); however, since the data has already passed the initial sediment chemistry screening step, observed toxicity in reference sediments is unlikely to be the result of significant chemical exposures.

Results

The Jamaica Bay reference dataset consisted of 45 acceptable samples with co-located benthic community, sediment toxicity, and sediment chemistry data. No extreme values were removed from the Jamaica Bay dataset because sediment toxicity was highly variable in samples that passed the initial chemical screen. Table 2-11 presents a summary of the Jamaica Bay dataset using the quantitative analysis of uncertainty approach to define acceptable data from the Jamaica Bay estuary that could be used as reference conditions for the LPRSA.

Table 2-11. Summary of benthic community metrics in the Jamaica Bay estuary reference dataset for the quantitative analysis of uncertainty

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Sample size	45	45	45	45	48
Minimum	110	3.5	0.519	0.232	1.0

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)
Maximum	282,150	48	2.76	0.875	8.0
Mean	27,395	23	1.66	0.602	3.6
Standard deviation	48,227	12	0.501	0.150	1.6
5th percentile[a]	516	7.1	0.820	0.399	1.5
10th percentile	675	8.2	0.916	0.424	1.9
25th percentile	2,875	14	1.44	0.485	2.5
Median	14,050	22	1.66	0.608	3.3
75th percentile	25,113	29	2.00	0.697	4.5
90th percentile	64,943	42	2.23	0.835	6.0
95th percentile	90,125	44	2.42	0.855	6.7

Source: USEPA REMAP (2002b)

Note: Three samples (i.e., JB310, JB315, and JB366) were removed from the Jamaica Bay dataset based on best professional judgment regarding low *A. abdita* survival values. Survival in sediments from those Jamaica Bay locations ranged from 1.1 to 11% of the negative control. No values were identified as extreme using the three-times-IQR approach described above.

[a] Reference threshold for benthic invertebrate community LOE.

IQR – interquartile range
LOE – line of evidence
REMAP – Regional Environmental Monitoring and Assessment Program
SDI – Swartz's dominance index
USEPA – US Environmental Protection Agency

The urban freshwater reference data from above Dundee Dam that were deemed acceptable for the quantitative analysis of uncertainty consisted of 15 samples (an increase from 5 samples, described in Table 2-6). No extreme values were identified or removed based on sediment toxicity. Table 2-12 presents the summary of those data from above Dundee Dam.

Table 2-12. Summary of benthic community metrics from the reference area above Dundee Dam for the quantitative analysis of uncertainty

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI[a]
Sample size	15	15	15	15	15	15
Minimum	2,070	13.7	1.76	0.617	3.67	6.84
Maximum	19,500	32.3	2.73	0.860	9.33	9.41
Mean	10,665	21.8	2.18	0.714	5.51	8.18
Standard deviation	5,038	4.66	0.274	0.0653	1.62	0.680
5th percentile	3,064	15.8	1.78	0.617	3.90	6.97
10th percentile	4,294	17.1	1.85	0.627	4.13	7.27
25th percentile	6,135	18.9	2.01	0.675	4.33	7.91
Median	11,500	21.3	2.16	0.720	5.00	8.25
75th percentile	14,150	24.7	2.31	0.739	6.17	8.55

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI[a]
90th percentile	15,740	26.3	2.55	0.783	7.67	8.89
95th percentile	17,120	28.4	2.67	0.811	8.63	9.08

Source: Windward (Draft)

[a] HBI was determined using NYSDEC tolerance values, with the exception of values for arthropods, which were taken from Hilsenhoff (1987).

HBI – Hilsenhoff Biotic Index

NYSDEC – New York State Department of Environmental Conservation

SDI – Swartz's dominance index

Windward – Windward Environmental LLC

The Mann-Whitney U test results based on the alternative datasets developed for the quantitative analysis of uncertainty are presented in Table 2-13 (and Appendix B, Tables B6-2c and B6-2d). There were similar significant differences between the LPRSA benthic invertebrate community metric datasets (for the uncertainty analysis) and reference area datasets.

Table 2-13. Results of the statistical analysis comparing the LPRSA benthic invertebrate community data with corresponding reference dataset for the quantitative analysis of uncertainty

LPRSA Benthic Salinity Zone	Area of Interest for Comparison	Is There a Significant Difference Between LPRSA Reference Area Data? (Mann-Whitney U test, alpha = 0.05)					
		Abundance (per m^2)[a]	Taxa Richness	Shannon-Weiner H'	Pielou's J'	SDI	HBI[b]
Upper estuary	Jamaica Bay (urban)	**yes; 4.14E-07**	**yes; 3.95E-08**	**yes; 9.45E-03**	no; 9.85E-01	**yes; 4.77E-02**	na[c]
Fluvial estuary	Jamaica Bay (urban)	**yes; 2.01E-03**	**yes; 4.24E-10**	**yes; 1.23E-08**	**yes; 1.16E-02**	**yes; 2.49E-08**	na[c]
Upper estuary	Mullica River and Great Bay (non-urban)	**yes; p = 1.28E-05**	**yes; p = 1.06E-05**	**yes; p = 2.30E-02**	no; p = 6.69E-01	no; p = 6.27E-02	na[c]
Fluvial estuary	Mullica River and Great Bay (non-urban)	no; p = 2.35E-01	**yes; p = 7.97E-06**	**yes; p = 1.40E-04**	**yes; p = 4.26E-02**	**yes; p = 7.95E-05**	na[c]
Tidal freshwater	Above Dundee Dam (urban)	no; p = 4.37E-01	**yes; p = 3.36E-04**	**yes; p = 4.70E-06**	no; p = 1.73E-05	**yes; p = 1.53E-05**	**yes; p = 2.77E-04**

Note: **Bold** text indicates significant result ($p < 0.05$); significance/p-values are based on lower-tailed Mann-Whitney U tests, unless otherwise noted.

[a] Abundance tested using two-tailed Mann-Whitney U test.

[b] HBI tested using upper-tailed Mann-Whitney U test.

[c] HBI is specific to freshwater, so comparison to Jamaica Bay or Mullica River/Great Bay data was not appropriate

HBI – Hilsenhoff Biotic Index

LPRSA – Lower Passaic River Study Area

na – not applicable

SDI – Swartz's dominance index

Using the 5th and/or 95th percentile reference values to establish location-by-location differences between LPRSA benthic metrics values and reference conditions, it was determined that data from several LPRSA SQT locations were outside the reference envelope (Table 2-14; Figure 2-9; Appendix B, Tables B3 and B4). Richness and diversity were the most consistently different metrics across salinity zones, whereas evenness and dominance were different at the fewest locations (at least in the upper and fluvial estuarine zones). In the upper estuarine zone, abundance was also frequently outside the reference envelope, while evenness and dominance were inside. In the fluvial estuarine zone, abundance was less frequently outside the reference envelope (than in the upper estuarine zone), but diversity, evenness, and dominance increased (relative to the upper estuarine zone). In the tidal freshwater zone, most of the benthic metrics (excepting abundance) were outside the reference envelope at a majority of LPRSA locations. More LPRSA locations exceeded the non-urban reference envelope than the urban reference envelope, as expected (Table 2-14).

Table 2-14. Summary of comparison of benthic community metrics within reference envelope for the quantitative analysis of uncertainty

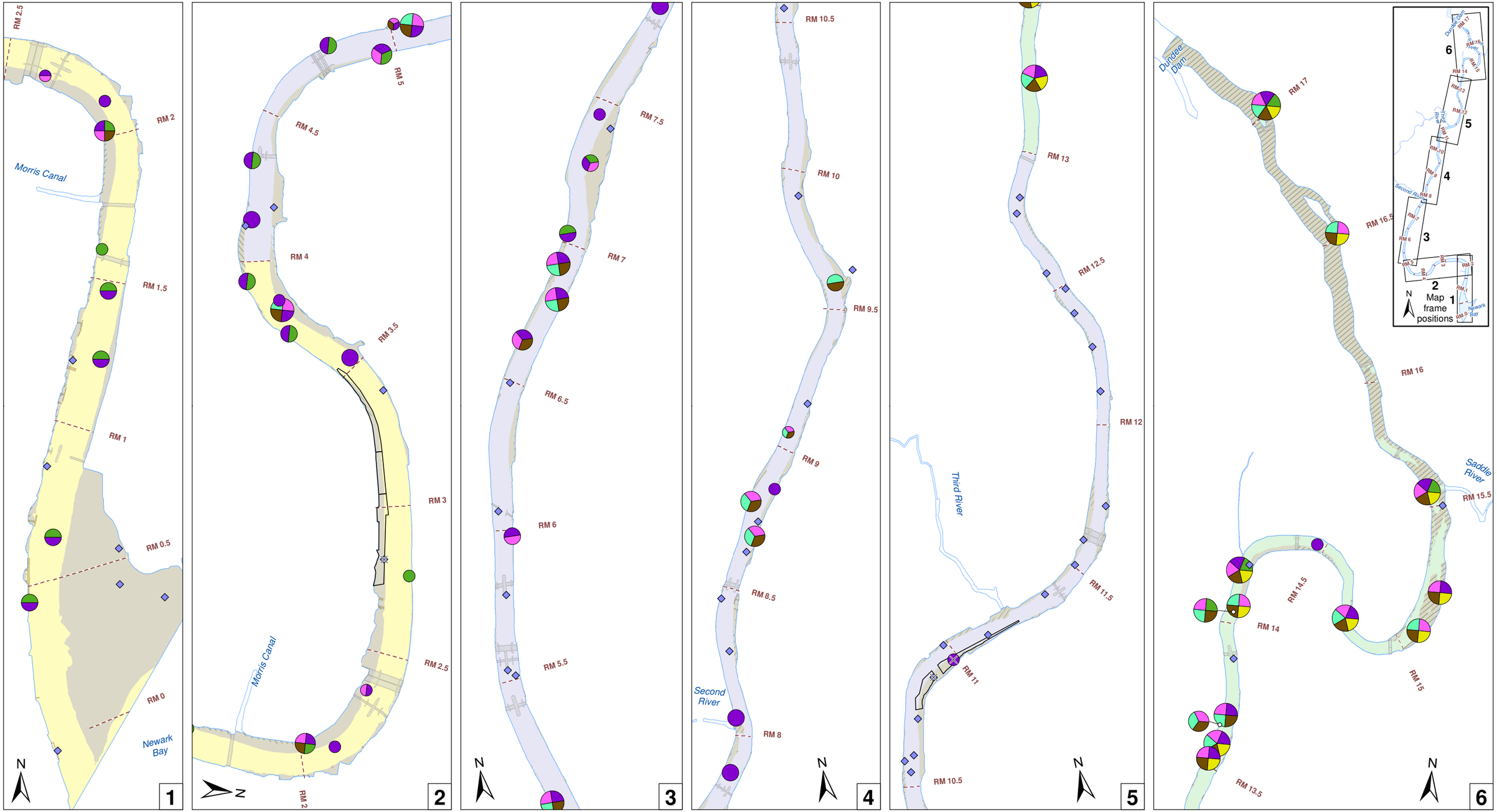
Benthic Salinity Zone	Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner H′	Pielou's J′	SDI (No. of Taxa)	HBI[a]
Comparison with Jamaica Bay (urban)							
Upper estuary (RM 0 to RM 4)	sample size	25	25	25	25	25	na[b]
	no. of locations outside reference envelope	9	12	3	1	2	na[b]
	percent of locations outside reference envelope	36%	48%	12%	4%	8%	na[b]
Fluvial estuary (RM 4 to RM 13)[c]	sample size	54	54	54	54	54	na[b]
	no. of locations outside reference envelope	5	17	11	7	9	na[b]
	percent of locations outside reference envelope	9%	31%	20%	13%	17%	na[b]
Comparison with Mullica River and Great Bay (non-urban)							
Upper estuary (RM 0 to RM 4)	sample size	25	25	25	25	25	na[b]
	no. of locations outside reference envelope	21	19	3	1	2	na[b]
	percent of locations outside reference envelope	84%	76%	12%	4%	8%	na[b]
Fluvial estuary (RM 4 to RM 13)[c]	sample size	54	54	54	54	54	na[b]
	no. of locations outside reference envelope	15	30	18	5	18	na[b]
	percent of locations outside reference envelope	28%	56%	33%	9%	33%	na[b]
Comparison with the area above Dundee Dam (urban)							
Tidal freshwater (RM 13 to RM 17.4)	sample size	18	18	18	18	18	18
	no. of locations outside reference envelope	4	10	14	10	14	11
	percent of locations outside reference envelope	22%	56%	78%	56%	78%	61%

[a] HBI was determined using NYSDEC tolerance values. HBI was the only metric compared to the maximum; all others were compared to the minimum.

[b] HBI applies to only tidal freshwater locations.

[c] Methods of sample collection differed between these two datasets (e.g., sieve sizes used were 1 mm or 0.5 mm); within the fluvial estuarine zone of the LPRSA, only samples upstream of RM 8.5 differed in this way.

HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River Study Area
na – not applicable
NYSDEC – New York State Department of Environmental Conservation
RM – river mile
SDI – Swartz's dominance index



Metric outside reference envelope[a]

- Abundance (per m^2)
- Taxa Richness
- Shannon-Wiener H' (log e)
- Pielou's J'
- Swartz's Dominance
- Hilsenhoff Biotic Index

- All metrics within reference envelope
- Remediated location[b]
- Dredge zone
- Upper Estuarine
- Fluvial Estuarine
- Tidal Freshwater

- Bridge
- Abutment
- Dock
- River mile
- Mudflat
- Gravel with fines

0 0.25 0.5 Miles

0 0.25 0.5 Kilometers

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

[a]Hilsenhoff Biotic Index value above the 95th percentile of the reference data; Abundance value above the 95th percentile or below the 5th percentile; all other metrics below the 5th percentile. Note: Locations downstream of RM 13 were compared to urban reference data from Jamaica Bay; locations upstream of RM 13 were compared to urban reference data from the area of the Passaic River above Dundee Dam.

[b]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

Windward LLC environmental

Figure 2-9. Benthic community metrics in the LPRSA compared with reference data representing urban habitats for the quantitative analysis of uncertainty
Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6370_Benthic metrics_Urban -alternative-_LSM_20160721.mxd

These results suggest somewhat depressed communities throughout the LPRSA, particularly in the tidal freshwater zone. Communities in the fluvial estuarine zone are exposed to a more stressful environment caused by daily fluctuations in overlying water chemistry (e.g., salinity). Salinity conditions are not expected to be as extreme in the Jamaica Bay estuary based on measured salinities in that areas (Appendix L, Attachment L3, Figure 2); salinities in Mullica River/Great Bay fluctuate to a greater extent, although they generally remain above interstitial salinities measured throughout the LPRSA (at 2009 SQT sampling locations).

Benthic communities in the LPRSA tidal freshwater zone are generally depressed relative to communities in the area above Dundee Dam (e.g., ≥ 56% of tidal freshwater SQT locations are outside the reference envelope for all benthic metrics, excepting abundance [22%]). This may be due, in part, to somewhat different habitats evaluated in the two areas; freshwater LPRSA locations tended to be shallower and to have coarser substrates than locations above Dundee Dam. The results may also be due, in part, to greater chemical contamination within the LPRSA relative to reference data screened for sediment chemistry (i.e., chemical impacts).

Metrics in the upper estuarine zone appear to be generally similar to urban reference conditions, with a maximum (across metrics) of 31% of LPRSA locations exceeding the reference condition for taxa richness.

2.4 SUMMARY

A comparison of LPRSA benthic community metrics with the reference datasets indicates that the benthic community in the upper estuarine salinity zone is somewhat different than the community in a less contaminated, urban estuary (Jamaica Bay), or in the non-urban estuary of Mullica River and Great Bay (Table 2-7). Most of the benthic community metrics for the upper estuarine zone and fluvial estuarine zone are significantly different than those in Jamaica Bay and Mullica River/Great Bay; the differences are substantiated somewhat by the location-by-location comparison of community metrics to reference envelope thresholds (Table 2-8). Specifically, there are notable differences in abundance and richness between the locations in the upper estuarine zone of the LPRSA and locations in Jamaica Bay. However, the analysis also suggests that there are inconsistent impacts across multiple metrics (e.g., 8, 4, and 4% exceedance rates for Shannon-Wiener H′, Pielou's J′, and SDI, respectively) (Table 2-8).

Although as a group most of the benthic community metrics for the fluvial estuarine zone are significantly different than those for Jamaica Bay or Mullica River/Great Bay (Table 2-7), the location-by-location comparison of community metrics to the estuarine reference envelope (Table 2-8) indicates that location-specific differences are sometimes minor and inconsistent across multiple metrics. For example, exceedance frequency of the urban reference condition in the LPRSA fluvial estuarine zone was ≤ 17% for abundance, Pielou's J′, and SDI (Table 2-8).

Based on a comparison of the LPRSA benthic metrics to those from above Dundee Dam, it appears that benthic communities in the tidal freshwater LPRSA are frequently impaired (Table 2-8). This is corroborated by Mann-Whitney U test results used to discern significant overall differences (Table 2-7). Only abundance was generally similar between the LPRSA and the area above Dundee Dam, exceeding the reference condition at only 22% of tidal freshwater LPRSA locations.

3 Sediment Toxicity Testing Line of Evidence

Sediment toxicity tests assess the ability of sediment to support benthic communities that include sensitive taxa. The results of toxicity tests conducted using LPRSA surface sediment samples were evaluated as an independent LOE in assessing risks to the benthic community (Appendix B, Tables B3, B4, and B5). Toxicity test data were also incorporated into the SQT and WOE analyses (Section 6.1 of the BERA main text; Appendix B, Tables B8 and B9).

Laboratory toxicity testing was conducted using sediment collected from 98 locations in the LPRSA.[11] LPRSA test results were compared with toxicity test results from USEPA-approved reference areas: Jamaica Bay and the Mullica River/Great Bay estuary, both of which represent estuarine reference conditions (and urban and non-urban conditions, respectively), and the area above Dundee Dam within the Passaic River (upstream of the LPRSA), which represents urban, freshwater reference conditions (Appendix B, Tables B3 and B4). LPRSA data were also compared to laboratory negative control results (Appendix B, Table B5). The reference area data are assumed to represent toxicological conditions in water bodies similar to the LPRSA, but without site-specific, regulated hazardous materials, but the Mullica River/Great Bay data do not represent urban conditions. The comparison of LPRSA data to reference area data provides insight into the potential impact of site-specific chemistry on sediment toxicity. The comparison of LPRSA data to negative control results is a quality control measure; negative controls are generally used to confirm the health of laboratory test organism cultures.[12]

3.1 METHODS

Sediment for toxicity testing was co-located with sediment that underwent chemical analysis (and benthic invertebrate community analysis). The samples were collected to a depth of 15 cm (6 in.) using a grab sampler. Replicate sediment samples ($n \geq 3$) from a given location were composited and homogenized.[13] A subsample was apportioned for toxicity testing, and the remainder of the sample was apportioned for chemical analysis.[14]

Two sediment toxicity tests were conducted using each of the 98 sediment samples according to the following methods:

[11] Benthic community survey data were not collected from 1 of 98 LPRSA sampling locations (LPRT16B) where sediment was collected in fall 2009 for sediment toxicity testing and chemical analysis.

[12] For example, test acceptability criteria for negative control survival are established in standard methods; failure of a negative control to meet the test acceptability criterion results in invalid test results.

[13] Additional information regarding the 2009 LPRSA sediment sampling event is provided in data reports (Windward 2014, 2018b).

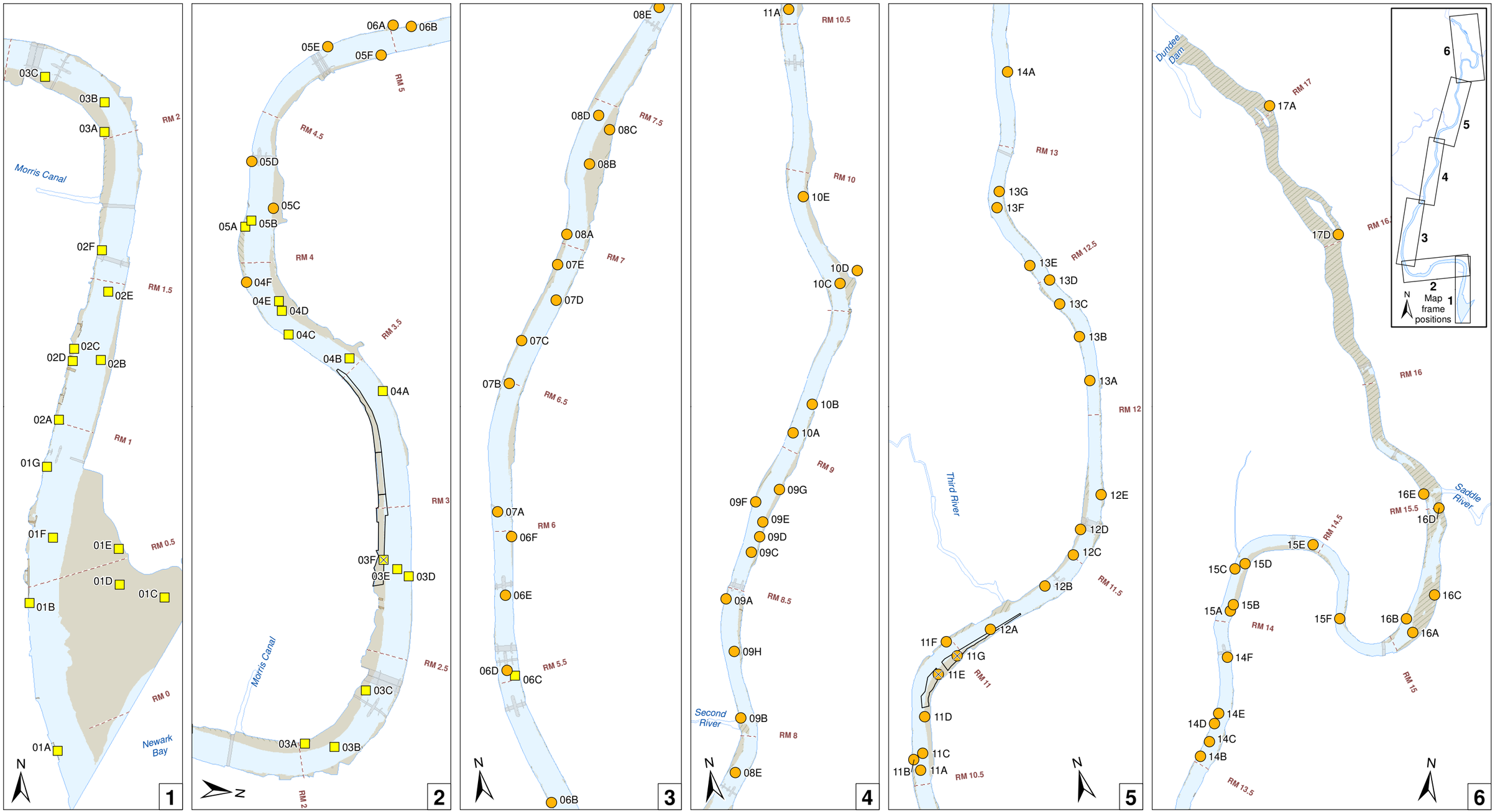
[14] Sediment samples for benthic community analysis were not homogenized prior to processing.

- Estuarine sediment samples
  - 10-day *A. abdita* survival test (ASTM 2008), with static renewal conditions
  - 28-day *H. azteca* survival and growth test using amphipods acclimated to 10 ppt salinity and conducted using sediment adjusted to 10 ppt salinity[15] (USEPA 2000)
- Freshwater sediment samples
  - 10-day *C. dilutus* survival and growth test (USEPA 2000)
  - 28-day *H. azteca* survival and growth test (USEPA 2000)

Samples were considered estuarine or freshwater for the purpose of sediment toxicity testing based on the interstitial water salinity at the time of sample collection. Interstitial water salinity was measured in the sediment samples after arrival at the toxicity testing laboratory. Samples with interstitial water salinity < 5 ppt were considered freshwater.[16] Samples with interstitial water salinity ≥ 5 ppt were considered estuarine. Based on these salinity measurements, 27 locations were considered estuarine, and 71 were considered freshwater (Figure 3-1). This definition of salinity differs from the salinity zones described in Section 2.1, which were based on a general understanding of the LPRSA as described in Section 2.2.1 of the BERA main text.

[15] Prior to testing, *H. azteca* were acclimated to 10 ppt seawater, which is within the known salinity tolerance of *H. azteca* (USEPA 2000).

[16] The threshold of 5 ppt was based on American Society for Testing and Materials (ASTM) (2007), which states that chironomid larvae are found in the field at a conductivity that ranges between 100 and 4,000 µS/cm (4,000 µS/cm at 0°C is equivalent to 4.1 ppt salinity).



Windward environmental LLC

0 0.25 0.5 Miles

0 0.25 0.5 0.75 Kilometers

- Freshwater
- Estuarine
- Freshwater remediated location[a]
- Estuarine remediated location[a]
- Dredge zone
- Bridge
- Abutment
- Dock
- River mile
- Mudflat
- Gravel with fines
- LPRSA

[a]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

Note: Locations with interstitial salinity ≤5 ppt were tested using freshwater methods, locations with interstitial salinity >5 ppt were tested using estuarine methods.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-1. Distribution of freshwater and estuarine benthic sediment toxicity testing locations in the LPRSA

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\5772_Benthic and tox sampling locations_LSM_20131206.mxd

Details of the toxicity test methods and results are presented in the *Fall 2009 Sediment Toxicity Test Data Report of the Lower Passaic River Study Area* (Windward 2018b). The toxicity tests were performed according to American Society for Testing and Materials (ASTM) and USEPA methods (ASTM 2007, 2008; USEPA 2000), which incorporate standard QA/QC procedures for evaluating the validity of test results. These procedures included the use of negative and positive controls and the periodic measurement of water quality during testing. Sediment samples were tested in batches to minimize the time between sample collection and testing, and to ensure that sample holding times were met for all samples. Each test batch was conducted with a batch-specific negative control, which was required to meet test-specific acceptability criteria. The batch-specific negative control was also used to calculate control-normalized toxicity values; control normalization was calculated by dividing the mean toxicity test result from an LPRSA location by the mean batch-specific negative control toxicity test value.[17] A third-party validator conducted a 100% validation of the toxicity test data, which included an initial evaluation of all data for completeness and accuracy, followed by a final evaluation of the overall quality and usability of the data (Dinnel 2010). The various toxicity tests were deemed acceptable based on the acceptable performance of negative controls in all batches.

CPG prepared a project-specific standard operation procedure (SOP) for the 28-day *H. azteca* survival and growth test that described the steps for acclimating and culturing the amphipods to water with 10 ppt salinity, and for conducting the sediment toxicity test exposure with overlying water adjusted to 10 ppt salinity. The USEPA-approved, project-specific SOP (Windward 2009) was reviewed by Dr. Chris Ingersoll, an aquatic toxicologist and branch chief of the US Geological Survey (USGS) Columbia Environmental Research Center, Columbia, Missouri, who has led the development of numerous toxicity test methods.

The *A. abdita* survival test method was modified from the method presented by ASTM (2008), and the change had the potential to alter toxicity test results. Specifically, the test was run using a static renewal exposure rather than a static exposure.[18] This could influence the test in two ways. Gently adding overlying water has the effect of stabilizing water quality conditions (e.g., dissolved oxygen [DO], ammonia) in test chambers, which should result in greater survival and growth. This does not constitute a positive bias because sediment toxicity is not altered; rather, the stressful conditions that are created as a result of the toxicity test itself are ameliorated. Static renewal can introduce positive bias into toxicity tests that are driven by volatile compounds. Small perturbations of overlying test chamber waters can increase volatilization and thereby decrease chemical exposures. In the LPRSA, key chemicals

[17] Control normalization is used to standardize toxicity test results to a constant level of laboratory test organism culture health.

[18] This deviation from standard method was suggested by USEPA Region 2 during the development of the USEPA-approved project-specific SOP for sediment toxicity testing (Windward 2009).

of interest (COIs) have low or no volatility under standard toxicity test conditions (e.g., temperature and atmospheric pressure), so the influence of renewal is expected to have had a negligible effect on *A. abdita* toxicity test exposures.

The comparison of LPRSA sediment toxicity data to reference area data was similar to the comparison of benthic invertebrate community metric data (Section 2; Appendix B, Tables B3 and B4), although in addition to the comparison to reference conditions, negative control data (from sediment toxicity tests) were compared with LPRSA data (Appendix B, Table B5). Statistical tests were used to identify significant differences between LPRSA data and negative control data. In order to select the most appropriate t-test for comparing negative controls to LPRSA data, two tests of initial assumptions were run: Levene's test of equality of variance (alpha = 0.05), and the Shapiro-Wilk test of normality (alpha = 0.05). If the data satisfied normality (Shapiro-Wilk $p \geq 0.05$) and equality of variance (Levene's $p < 0.05$), then the standard Student's t-test was applied (one-tailed, alpha = 0.05). If variances were not equal between the LPRSA data and negative control, then the Welch's test was applied (one-tailed, alpha = 0.05). If normality was not satisfied, then toxicity test results were converted to rankit units[19] and tested using either the Student's t-test or Welch's test (one-tailed and alpha = 0.05 for either test), depending on the equality of variance between LPRSA and negative control rankit values. One-tailed tests were used to identify whether LPRSA sediment toxicity test results were significantly lower (or more toxic) than the negative control.

3.2 RESULTS

This section presents the results of the toxicity testing of sediment collected in the LPRSA in fall 2009 (Windward 2018b). Table 3-1 provides the control-normalized toxicity data for the LPRSA estuarine and freshwater samples. The complete dataset (e.g., including raw toxicity values and sample standard deviations) for the toxicity test results is provided in Appendix K, Table K5.

Table 3-1. LPRSA sediment toxicity data

Location ID	Test Type	*C. dilutus* Survival (% of Control)	*C. dilutus* Biomass (% of Control)[a]	*H. azteca* Survival (% of Control)	*H. azteca* Biomass (% of Control)[b]	*A. abdita* Survival (% of Control)
LPRT01A	estuarine	nt	nt	87.9	82.3	89.7
LPRT01B	estuarine	nt	nt	89.2	71.5	97.3
LPRT01C	estuarine	nt	nt	41.1	25.9	98.6
LPRT01D	estuarine	nt	nt	81.1	69.9	59.5
LPRT01E	estuarine	nt	nt	24.6	18.6	90.9
LPRT01F	estuarine	nt	nt	94.6	69.0	87.6
LPRT01G	estuarine	nt	nt	89.2	64.7	99.5

[19] Rankits were calculated using the `qnorm` function in R (R Core Team 2015).

Table 3-1. LPRSA sediment toxicity data

Location ID	Test Type	*C. dilutus* Survival (% of Control)	*C. dilutus* Biomass (% of Control)[a]	*H. azteca* Survival (% of Control)	*H. azteca* Biomass (% of Control)[b]	*A. abdita* Survival (% of Control)
LPRT02A	estuarine	nt	nt	86.5	55.6	91.9
LPRT02B	estuarine	nt	nt	71.7	59.6	71.4
LPRT02C	estuarine	nt	nt	94.5	67.9	98.6
LPRT02D	estuarine	nt	nt	82.1	56.7	101.9
LPRT02E	estuarine	nt	nt	82.5	68.3	80.0
LPRT02F	estuarine	nt	nt	90.6	60.6	85.4
LPRT03A	estuarine	nt	nt	82.5	84.4	62.7
LPRT03B	estuarine	nt	nt	87.9	83.0	96.2
LPRT03C	estuarine	nt	nt	56.2	42.6	93.1
LPRT03D	estuarine	nt	nt	85.2	64.2	89.7
LPRT03E	estuarine	nt	nt	86.5	95.1	91.9
LPRT03F	estuarine	nt	nt	72.6	43.2	89.8
LPRT04A	estuarine	nt	nt	98.7	95.9	84.3
LPRT04B	estuarine	nt	nt	87.9	70.8	23.8
LPRT04C	estuarine	nt	nt	35.6	17.0	16.4
LPRT04D	estuarine	nt	nt	68.5	51.4	76.7
LPRT04E	estuarine	nt	nt	61.7	35.3	88.7
LPRT04F	freshwater	82.9	82.1	86.5	44.0	nt
LPRT05A	estuarine	nt	nt	6.8	2.6	39.4
LPRT05B	estuarine	nt	nt	16.4	6.0	94.2
LPRT05C	freshwater	95.8	70.3	52.8	18.6	nt
LPRT05D	freshwater	88.6	71.3	67.6	31.5	nt
LPRT05E	freshwater	95.9	70.6	70.2	45.0	nt
LPRT05F	freshwater	104.3	65.1	33.8	16.0	nt
LPRT06A	freshwater	94.3	59	87.9	38.5	nt
LPRT06B	freshwater	82.9	65.6	64.9	21.3	nt
LPRT06C	estuarine	nt	nt	52.0	25.5	94.2
LPRT06D	freshwater	90.1	62.1	63.6	19.3	nt
LPRT06E	freshwater	90.1	73.8	81.1	37.0	nt
LPRT06F	freshwater	88.0	60.7	89.5	58.2	nt
LPRT07A	freshwater	95.9	90	67.2	30.7	nt
LPRT07B	freshwater	95.8	64.6	21.6	7.3	nt
LPRT07C	freshwater	100.0	52.2	71.6	36.1	nt
LPRT07D	freshwater	89.3	74.6	65.6	35.0	nt

Table 3-1. LPRSA sediment toxicity data

Location ID	Test Type	*C. dilutus* Survival (% of Control)	*C. dilutus* Biomass (% of Control)[a]	*H. azteca* Survival (% of Control)	*H. azteca* Biomass (% of Control)[b]	*A. abdita* Survival (% of Control)
LPRT07E	freshwater	85.3	59.2	74.6	41.6	nt
LPRT08A	freshwater	97.3	75.6	50.7	21.8	nt
LPRT08B	freshwater	100.0	68.2	62.6	39.9	nt
LPRT08C	freshwater	101.3	59.7	35.8	22.5	nt
LPRT08D	freshwater	72.9	83.6	77.1	42.9	nt
LPRT08E	freshwater	105.7	94.4	69.0	46.9	nt
LPRT09A	freshwater	94.3	57.9	36.5	12.5	nt
LPRT09B	freshwater	85.6	80.2	70.6	72.6	nt
LPRT09C	freshwater	88.2	69.2	67.6	60.2	nt
LPRT09D	freshwater	100.0	83.1	79.1	43.8	nt
LPRT09E	freshwater	97.3	75.6	59.7	32.8	nt
LPRT09F	freshwater	92.0	90.5	43.3	21.1	nt
LPRT09G	freshwater	93.3	97.5	103.0	63.5	nt
LPRT09H	freshwater	84.3	78.5	81.1	49.7	nt
LPRT10A	freshwater	78.9	34.5	64.7	41.4	nt
LPRT10B	freshwater	15.8	4.7	19.2	9.0	nt
LPRT10C	freshwater	98.7	36.2	69.2	52.4	nt
LPRT10D	freshwater	80.3	57.3	101.5	92.0	nt
LPRT10E	freshwater	98.6	72.6	37.4	25.5	nt
LPRT11A	freshwater	80.3	52.3	93.4	76.3	nt
LPRT11B	freshwater	47.4	20.3	64.7	37.4	nt
LPRT11C	freshwater	82.9	77.3	64.7	50.8	nt
LPRT11D	freshwater	88.2	55.8	79.4	60.7	nt
LPRT11E	freshwater	92.1	46.3	89.8	52.4	nt
LPRT11F	freshwater	94.7	45.8	52.9	42.8	nt
LPRT11G	freshwater	72.4	12.4	16.2	6.1	nt
LPRT12A	freshwater	94.7	55.9	67.6	32.1	nt
LPRT12B	freshwater	97.4	38	83.9	47.7	nt
LPRT12C	freshwater	80.3	29.3	58.8	27.0	nt
LPRT12D	freshwater	101.4	48.9	58.8	25.9	nt
LPRT12E	freshwater	90.8	25.4	52.9	27.3	nt
LPRT13A	freshwater	86.8	51.1	55.9	30.1	nt
LPRT13B	freshwater	93.5	52	90.9	48.3	nt
LPRT13C	freshwater	100.0	62.7	84.8	40.9	nt

Table 3-1. LPRSA sediment toxicity data

Location ID	Test Type	*C. dilutus* Survival (% of Control)	*C. dilutus* Biomass (% of Control)[a]	*H. azteca* Survival (% of Control)	*H. azteca* Biomass (% of Control)[b]	*A. abdita* Survival (% of Control)
LPRT13D	freshwater	93.5	38	92.7	51.8	nt
LPRT13E	freshwater	96.1	73.4	90.9	45.8	nt
LPRT13F	freshwater	94.8	103.7	81.8	53.7	nt
LPRT13G	freshwater	98.7	66	93.9	53.0	nt
LPRT14A	freshwater	92.1	44.5	44.1	27.3	nt
LPRT14B	freshwater	98.7	48.3	45.6	16.1	nt
LPRT14C	freshwater	92.1	61.3	85.3	52.3	nt
LPRT14D	freshwater	80.3	37.3	63.3	26.8	nt
LPRT14E	freshwater	96.1	50.7	52.9	21.4	nt
LPRT14F	freshwater	92.1	68.1	88.2	48.2	nt
LPRT15A	freshwater	100.0	54.1	58.8	35.9	nt
LPRT15B	freshwater	101.4	49.5	78.0	55.0	nt
LPRT15C	freshwater	93.5	43.1	80.9	46.0	nt
LPRT15D	freshwater	78.9	35.9	73.5	44.5	nt
LPRT15E	freshwater	97.4	46.4	89.8	59.4	nt
LPRT15F	freshwater	92.1	43.8	95.6	63.5	nt
LPRT16A	freshwater	73.7	53.8	78.5	54.2	nt
LPRT16B[c]	freshwater	78.9	66.3	95.6	114.5	nt
LPRT16C	freshwater	88.2	79.1	97.1	105.2	nt
LPRT16D	freshwater	96.1	101.2	86.8	63.4	nt
LPRT16E	freshwater	84.2	89.5	92.7	102.5	nt
LPRT17A	freshwater	78.9	43.4	98.6	64.9	nt
LPRT17D	freshwater	96.1	51.6	104.5	70.2	nt

Source: Windward (2018b)

[a] Biomass for *C. dilutus* was calculated as the total AFDW for each replicate divided by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.

[b] Biomass for *H. azteca* was calculated as the total weight for each replicate divided by the initial number of organisms introduced into the test chamber.

[c] LPRT16B was not analyzed using the full SQT analysis in the BERA because benthic community data were not collected at that sampling location.

AFDW – ash-free dry weight
BERA – baseline ecological risk assessment
ID – identification
LPRSA – Lower Passaic River Study Area
nt – not tested
SQT – sediment quality triad

For the 27 estuarine samples, the survival of *A. abdita* ranged from 15 to 97% of the negative control results, with a mean of 78% of control survival. *H. azteca* survival results were similar, ranging from 7 to 98% of the negative control, with a mean of

71% of control survival. *H. azteca* biomass ranged from 3 to 96% of the negative control with a mean of 55% of control results.

For the 71 freshwater samples, minimum survival values for both *C. dilutus* and *H. azteca* were 16% of the negative control; maximum survival was 110% of the control for *C. dilutus* and 100% of control for *H. azteca*. Mean survival values, as a percentage of the negative control, were 90 and 70% for *C. dilutus* and *H. azteca*, respectively.

3.2.1 Negative control data

3.2.1.1 Results

Detailed results from comparisons of LPRSA sediment toxicity test data to negative control data are provided in Appendix B, Table B5-1. Table 3-2 provides a summary of the comparisons between LPRSA and negative control toxicity test data.

Table 3-2. Summary of statistical tests comparing LPRSA sediment toxicity test results to negative control results

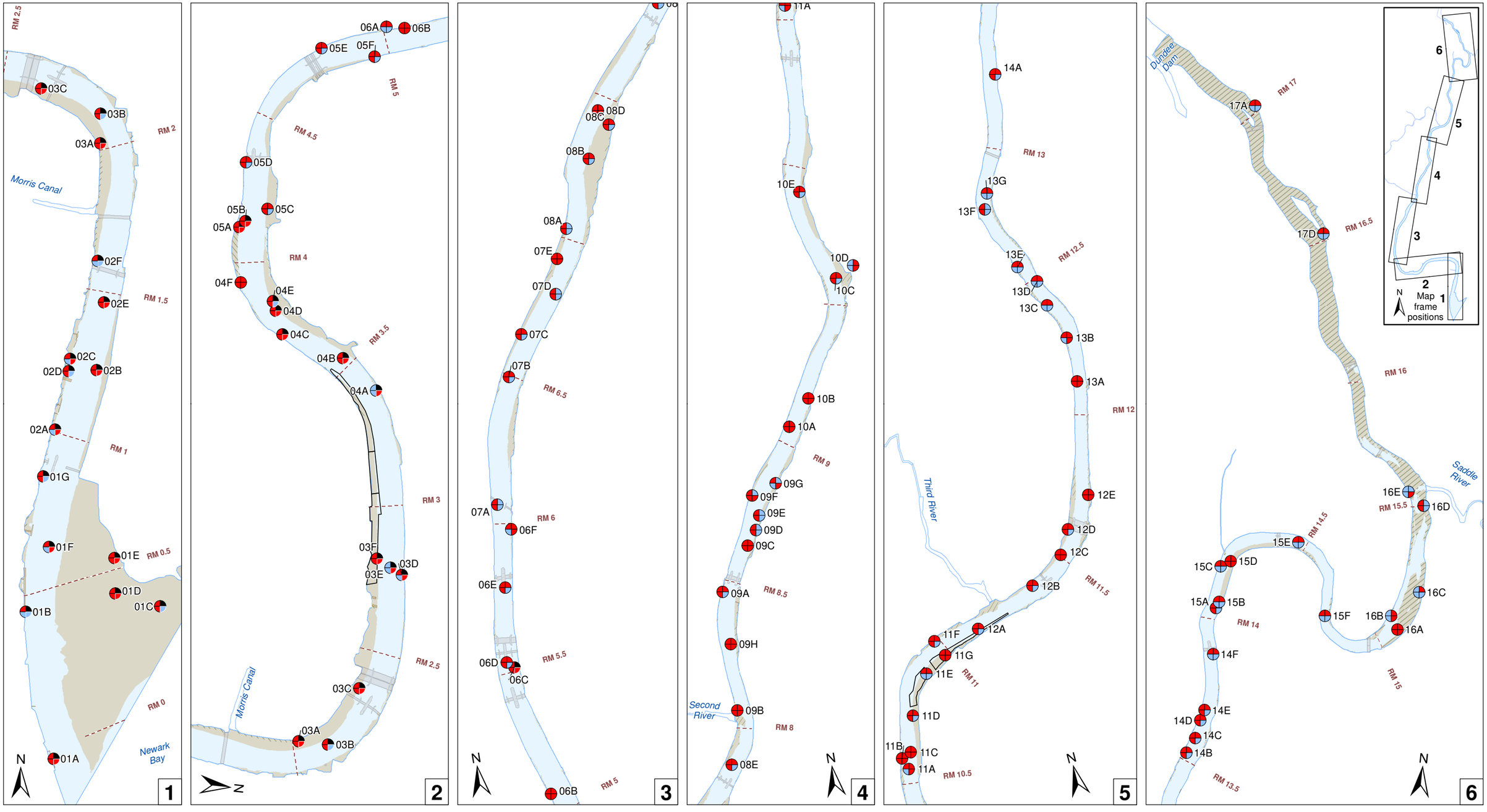
Toxicity Test Endpoint	No. Significantly Different LPRSA Locations[a]	No. of LPRSA Locations	Percentage Significantly Different LPRSA Locations[a]
C. dilutus 10-day biomass	61	71	86%
C. dilutus 10-day survival	25	71	35%
H. azteca 28-day biomass	92	98	94%
H. azteca 28-day survival	69	98	70%
A. abdita 10-day survival	20	27	74%

[a] Significance based on Student's or Welch's t-tests using raw or rankit data, as appropriate; Section 3.1 provides the decision process for selecting appropriate test methods.

LPRSA – Lower Passaic River Study Area

3.2.1.2 Summary of results

Based on the results presented in Table 3-2, it can be seen that a substantial number of LPRSA locations have sediment toxicity test results significantly different from negative control results. Figure 3-2 shows the distributions of negative control analyses outcomes in the LPRSA. Negative controls performed adequately, indicating that the laboratory cultures were healthy. Based on these results, it can be concluded that LPRSA sediments are toxic relative to clean, controlled sediment conditions (i.e., laboratory negative controls).



Toxicity test type

Estuarine	*Hyalella* Biomass	*Ampelisca* Biomass (no data)	
	Survival	Survival	
Freshwater	*Hyalella* Biomass	*Chironomus* Biomass	
	Survival	Survival	

T-test result
- Significant
- Not significant

× Remediated location[a]

Dredge zone

- River mile
- Bridge
- Abutment
- Dock
- Mudflat
- Gravel with fines
- LPRSA

0 0.25 0.5 Miles

0 0.25 0.5 0.75 Kilometers

[a]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-2. Results from comparisons of LPRSA toxicity test data to negative control results

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Prepared by mikey 5/30/2019, W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6450_Bioassay negative control results_LSM_20160721.mxd

Although this result provides some information regarding the magnitude of toxicity observed in LPRSA locations, it does not provide a practical means of comparison for natural sediment conditions. Negative control conditions are intended to ensure that toxicity test organisms are sufficiently healthy to perform adequately during toxicity testing. LPRSA sediment is different from negative control sediments in terms of both contamination by hazardous substances and the sediment matrix. Therefore, it is unclear whether significant differences from the negative control are caused by hazardous substances.

Comparison of LPRSA data to natural reference sediments (with relatively low contamination) (Section 3.3) provides another basis for comparison that partly controls for sediment chemistry and matrix effects.[20]

3.2.2 Reference area data

LPRSA sediment toxicity test results were compared with the toxicity test data from USEPA-approved reference areas to evaluate whether LPRSA sediment poses a potential risk to the benthic community (in excess of reference conditions). The reference area toxicity data represents a baseline level of toxicity for the LPRSA in the absence of the release of site-related hazardous substances. The use of reference conditions in evaluating LPRSA toxicity test data is consistent with USEPA guidance on the use of reference data in ERAs (USEPA 2002, 2005a).

Data from the three reference areas selected by USEPA (i.e., Jamaica Bay, Mullica River/Great Bay, and the area above Dundee Dam) were compiled for comparison with sediment toxicity test data from the LPRSA (Appendix B, Tables B3-1 and B4-1). Mullica River freshwater sediment toxicity data could not be used because they were generally not co-located with benthic invertebrate community data, or because they were based on a non-comparable test organism (i.e., *A. abdita*, which was not tested in the area above Dundee Dam). A sediment toxicity LOE analysis could therefore not be conducted for the LPRSA using Mullica River freshwater as a reference for comparison.

Although reference conditions are ideally non-toxic due to the relatively low level of historical chemical pollution in reference areas, there is the possibility for unforeseen toxicity in such areas. For this reason, reference area datasets were limited using sediment toxicity screening criteria (Section 2.3.1; Appendix B, Tables B3-3 and B4-3). Samples were also screened based on sediment chemical concentrations to eliminate the potential for contamination in accepted reference sediment samples.

[20] Physical and geochemical differences between LPRSA sediment and reference area sediment are still possible. There is no single "best" approach to evaluating sediment, which is why a WOE approach is generally taken to characterizing sediment quality. Additional context is presented in Appendix L (Attachment L3), which shows a comparison between reference area and LPRSA sediment chemistry and physical characteristic data.

3.2.2.1 Estuarine reference data

Sediment toxicity data from Jamaica Bay (representing urban estuarine habitat conditions) and the Mullica River/Great Bay estuary (representing non-urban estuarine habitat conditions) were compiled from regional datasets for comparison with LPRSA *A. abdita* toxicity test results (Appendix L, Table L8). Location selection, sample collection, and analyses were performed by others (USEPA 2011). Because regional reference data were not available to evaluate LPRSA *H. azteca* toxicity test results, toxicity data based on sediment samples collected by CPG in the freshwater area above Dundee Dam were used as a reference dataset for both survival and biomass of *H. azteca* in estuarine toxicity test locations in the LPRSA. The representativeness of data from above Dundee Dam as a reference for the estuarine LPRSA data is discussed in the uncertainty evaluation for the sediment toxicity LOE (Section 3.3.4).

Table 3-3 presents the summary of the Jamaica Bay and Mullica River/Great Bay reference area *A. abdita* toxicity data from acceptable reference locations (i.e., low chemistry and control-normalized survival ≥ 80%) (Appendix B, Tables B3-2 and B3-3). Survival of *A. abdita* in the 35 acceptable Jamaica Bay samples ranged from 80 to 108% of the negative control results, with a mean control-normalized survival of 95.4% (Table 3-3). Survival of *A. abdita* in the 12 acceptable Mullica River/Great Bay samples ranged from 84.4 to 107% of the negative control results, with a mean control-normalized survival of 95.9% (Table 3-3). Additional information regarding the screening of reference area datasets is presented in Appendix B (Table B3-3).

Table 3-3. Summary of the Jamaica Bay *A. abdita* survival data (fraction of control)

Statistic	Jamaica Bay	Mullica River and Great Bay
Sample size	35	12
Minimum	0.800	0.844
Maximum	1.08	1.07
Mean	0.945	0.959
Standard deviation	0.0617	0.069
5th percentile[a]	0.845	0.852
10th percentile	0.872	0.861
25th percentile	0.906	0.917
Median	0.950	0.984
75th percentile	0.990	1.00
90th percentile	1.01	1.01
95th percentile	1.03	1.04

Source: USEPA (2016); NOAA (2013)

[a] Reference envelope threshold for *A. abdita* survival.

NOAA – National Oceanic and Atmospheric Administration

USEPA – US Environmental Protection Agency

3.2.2.2 Freshwater reference data

CPG collected sediment samples within a 4-mile section of the Passaic River above Dundee Dam (representing a freshwater reference area) for comparison with LPRSA *H. azteca* and *C. dilutus* toxicity test results. Table 3-4 provides a summary of the toxicity data for acceptable locations from the area above Dundee Dam.

Table 3-4. Summary of toxicity reference data (fraction of control) from the area above Dundee Dam

Statistic	*C. dilutus*		*H. azteca*	
	Survival	Biomass[a]	Survival	Biomass[b]
Sample size	5	5	5	5
Minimum	0.82	0.775	0.76	0.355
Maximum	0.96	1.03	0.94	0.471
Mean	0.89	0.845	0.82	0.421
Standard deviation	0.053	0.103	0.075	0.048
5th percentile[a]	0.83	0.778	0.76	0.362
10th percentile	0.84	0.781	0.76	0.369
25th percentile	0.87	0.791	0.77	0.389
Median	0.90	0.813	0.81	0.434

Statistic	*C. dilutus*		*H. azteca*	
	Survival	Biomass[a]	Survival	Biomass[b]
75th percentile	0.91	0.819	0.83	0.454
90th percentile	0.94	0.944	0.90	0.464
95th percentile	0.95	0.986	0.92	0.468

Source: Windward (2018a)

[a] Biomass for *C. dilutus* was calculated as the total AFDW for each replicate divided by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.

[b] Biomass for *H. azteca* was calculated as the total weight for each replicate divided by the initial number of organisms introduced into the test chamber.

AFDW – ash-free dry weight

Windward – Windward Environmental LLC

3.2.3 Comparison with reference data

The LPRSA estuarine and freshwater toxicity test data were compared with reference area data following the steps outlined in Section 2.1 and shown in Figure 2-1. The Mann-Whitney U test was used to determine whether LPRSA toxicity data were significantly different from those in the reference areas (Appendix B, Table B6-2b), and a location-by-location analysis of differences was conducted by comparing toxicity data from LPRSA locations to the 5th percentile reference values (Appenix B, Tables B3-6, B3-8, and B4-5). Ultimately, comparisons between LPRSA toxicity data and 5th percentiles of reference area toxicity datasets were used to assign weight to the sediment toxicity LOE in the WOE analyses (Section 6.1 of the BERA main text). Section 2.3.1 provides additional rationale for the use of reference area data.

3.2.3.1 Mann-Whitney U Test

Results of the Mann-Whitney U test comparison of LPRSA sediment toxicity test data to reference area sediment toxicity datasets are provided in Table 3-5.

Table 3-5. Results of the statistical analysis comparing the LPRSA toxicity test data with reference data

Area of Interest	Toxicity Test Type	Is there a significant difference between LPRSA and reference area data? (Mann-Whitney U Test, alpha = 0.05)[a]				
		C. dilutus Survival (% of Control)	*C. dilutus* Biomass (% of Control)[b]	*H. azteca* Survival (% of Control)	*H. azteca* Biomass (% of Control)[c]	*A. abdita* Survival (% of Control)
Jamaica Bay	estuarine	nt	nt	nt	nt	**yes; p = 1.20E-06**
Mullica River and Great Bay	estuarine	nt	nt	nt	nt	no; p ~ 1.00
Above Dundee Dam	estuarine[d]	nt	nt	no; p = 4.08E-01	no; p = 9.27E-01	nt
Above Dundee Dam	freshwater	no; p = 3.97E-01	no; p = 5.22E-02	**yes; p = 2.32E-02**	no; p = 7.28E-01	nt

Note: Mann-Whitney U test conducted as a one-tailed test (alpha = 0.05). An "nt" value is reported for tests that were not conducted for the given test species in the given reference area and/or in the LPRSA.

Bold text indicates significant result ($p < 0.05$).

[a] Based on the Mann-Whitney U Test.

[b] Biomass for *H. azteca* was calculated as the total weight for each replicate divided by the initial number of organisms introduced into the test chamber.

[c] Biomass for *C. dilutus* was calculated as the total AFDW for each replicate divided by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.

[d] *H. azteca* test results from the estuarine LPRSA could not be compared to estuarine reference area data; instead, they were compared to freshwater *H. azteca* test results from above Dundee Dam.

AFDW – ash-free dry weight
LPRSA – Lower Passaic River Study Area
nt – not tested

Based on the results of Mann-Whitney U tests (Table 3-5), it appears that survival of *H. azteca* was significantly lower in freshwater LPRSA sediments than above Dundee Dam, and survival of *A. abdita* was significantly lower in estuarine LPRSA sediments than in Jamaica Bay. Biomass of *H. azteca* in acceptable reference samples from above Dundee Dam ranged from only 36 to 47% of the negative control, indicating a fairly substantial growth effect on this test species that was unrelated to LPRSA-specific chemical contamination.

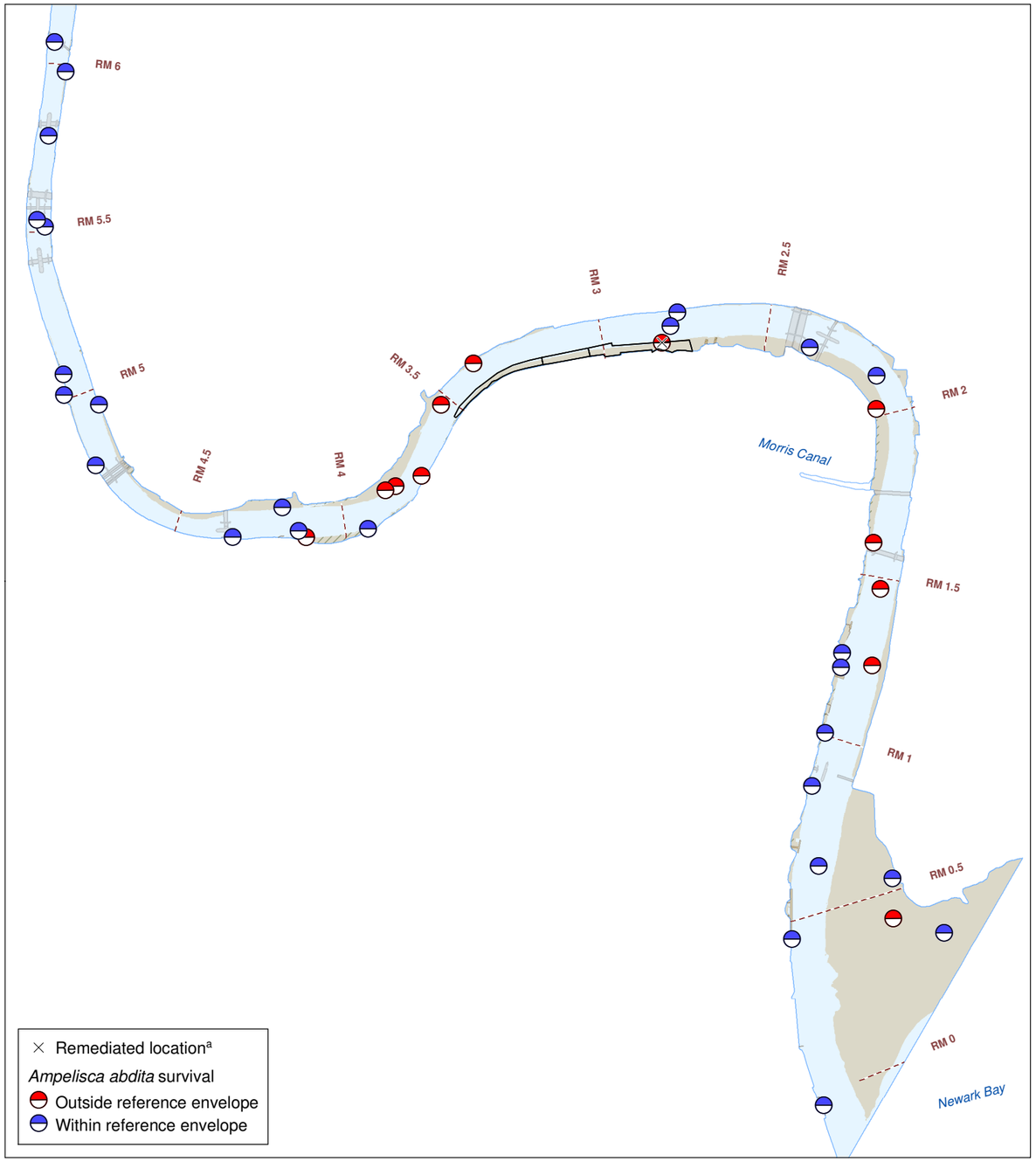
3.2.3.2 LPRSA locations below the reference area envelope

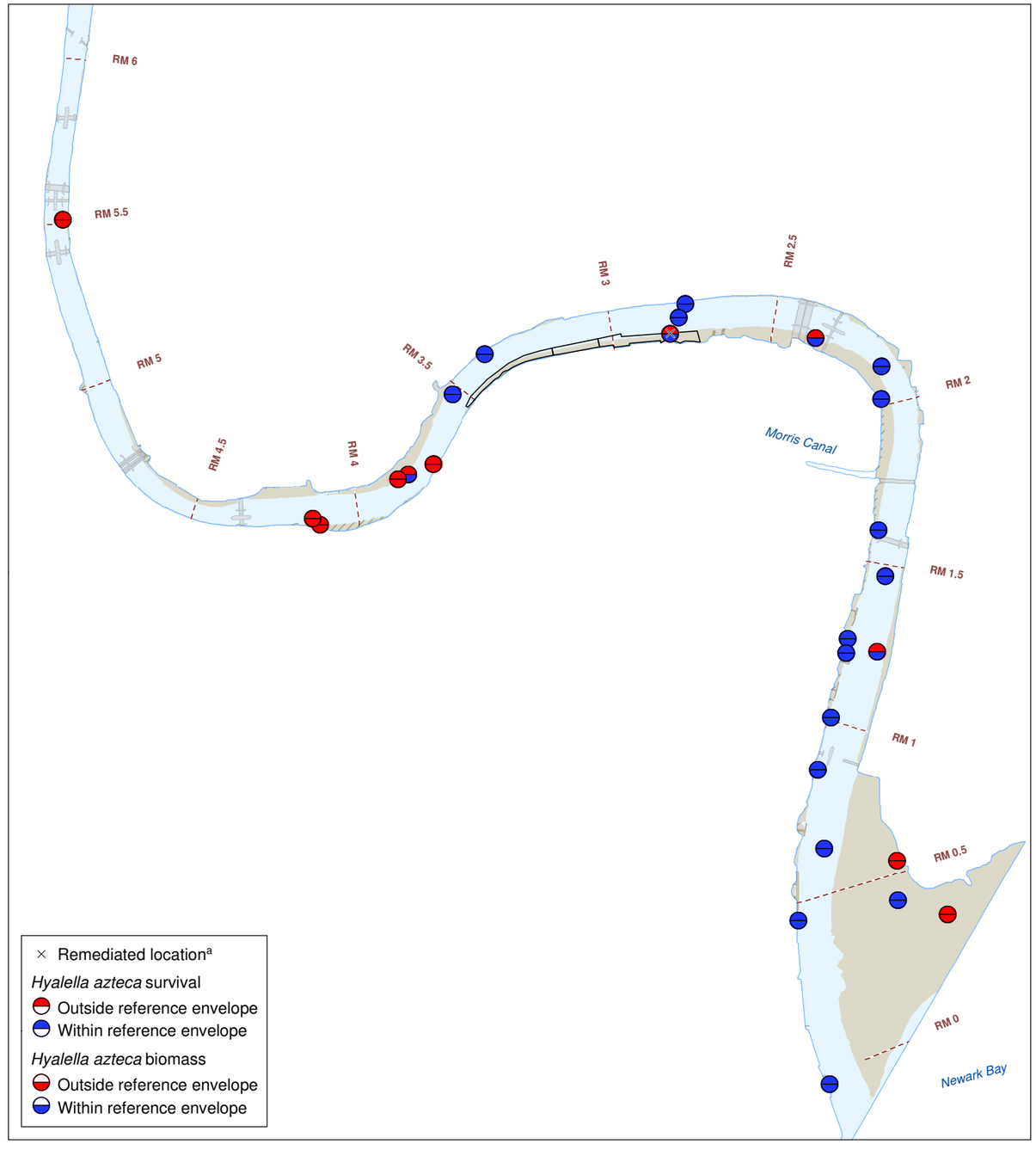
The result of location-by-location comparisons between LPRSA and reference area sediment toxicity data are summarized in Figures 3-3 to 3-6 and Table 3-6. Results are also provided in tabular form in Appendix B (Tables B3-6, B3-8, and B4-5).

Figures 3-3 and 3-6 present LPRSA locations outside the reference thresholds for *A. abdita* survival based on urban and non-urban reference datasets, respectively. Figure 3-3 presents LPRSA locations exceeding the reference thresholds for *H. azteca* survival and biomass from tests conducted for estuarine toxicity test locations (interstitial water salinity ≥ 5 ppt). Figure 3-4 presents LPRSA locations exceeding the

reference thresholds for *C. dilutus* survival and biomass. Figure 3-5 presents LPRSA locations exceeding the reference thresholds for *H. azteca* survival and biomass from tests conducted for freshwater toxicity test locations (interstitial water salinity < 5 ppt).

Table 3-6 provides a summary by toxicity test endpoint of LPRSA sediment toxicity test results below the 5th percentiles of reference values (i.e., reference envelope thresholds).





Windward environmental LLC

N

0 0.25 0.5 Mile

Dredge zone
Bridge
Abutment
Dock
River mile
Mudflat
Gravel with fines
LPRSA

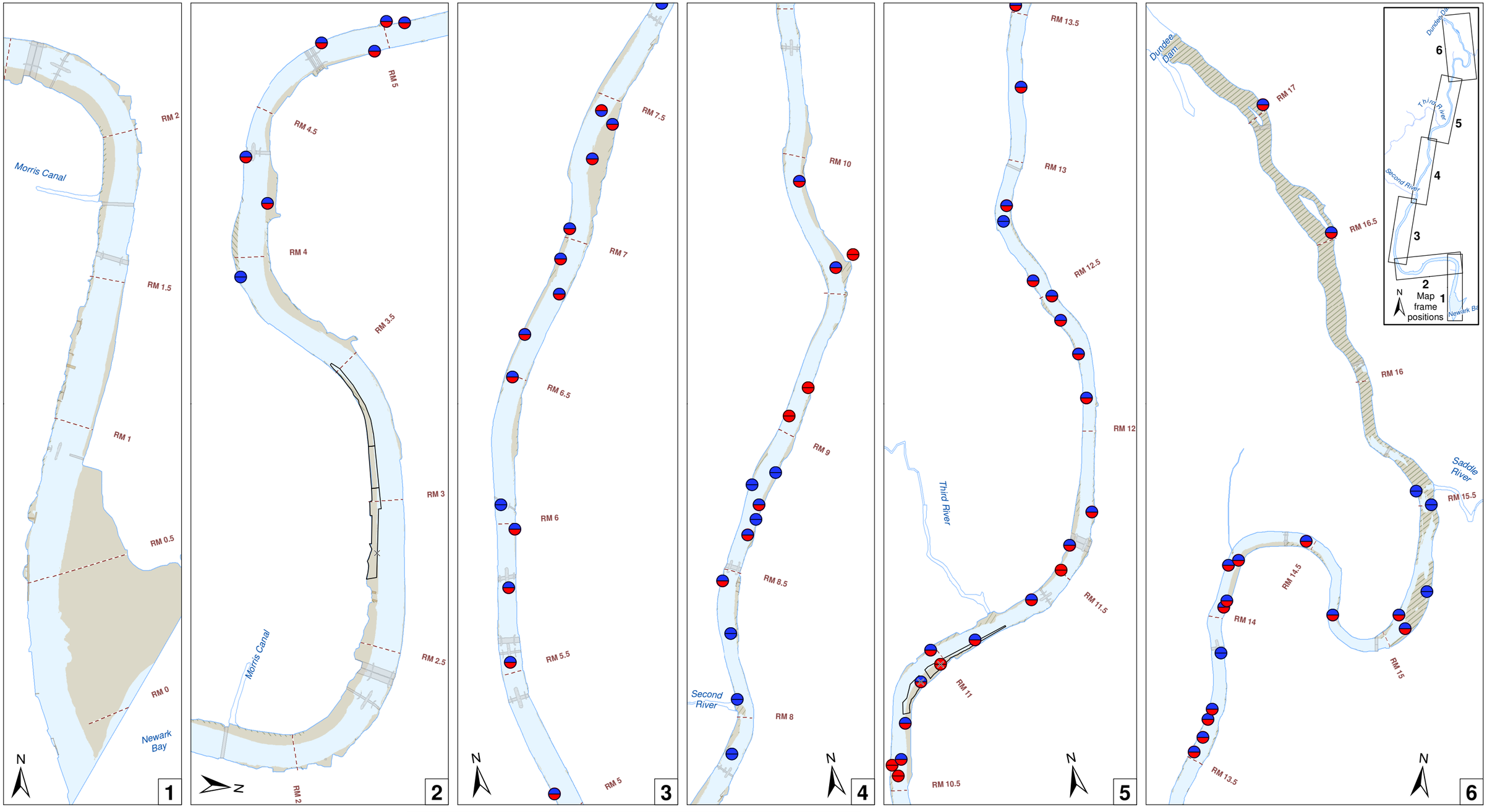
[a]One sample was collected in the Lister Ave. dredge area at RM 2.8

Note: LPRSA *A. abdita* toxicity data were compared to urban reference data from Jamaica Bay; LPRSA *H. azteca* toxicity data were compared to urban reference data from the area of the Passaic River above Dundee Dam.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-3. *Ampelisca abdita* survival and *Hyalella azteca* survival and biomass results from estuarine LPRSA locations compared with urban reference data
Appendix P. Lower Passaic River Study Area
Baseline Ecological Risk Assessment
FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6371_Ampelisca Hyalella urban estuarine results_LSM_20160721.mxd



Chironomus dilutus survival
- Outside reference envelope
- Within reference envelope

Chironomus dilutus biomass
- Outside reference envelope
- Within reference envelope

- Remediated location[a]
- Dredge zone
- Bridge
- Abutment
- Dock

- River mile
- Mudflat
- Gravel with fines
- LPRSA

0 0.25 0.5 Miles

0 0.25 0.5 0.75 Kilometers

[a]Two samples were collected in the RM 10.9 dredge area.

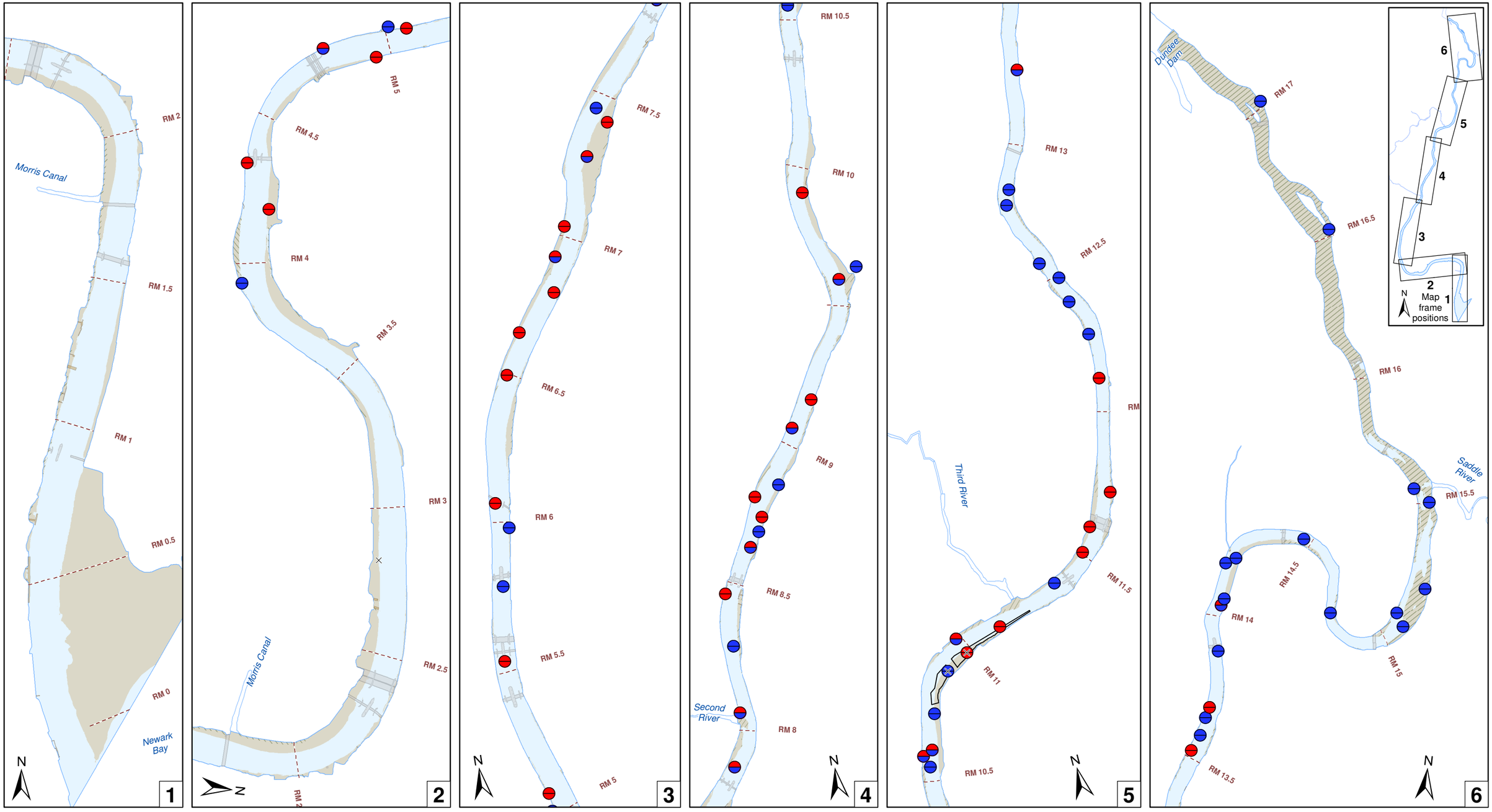
Note: Data are control-normalized; reference data are from the area of the Passaic River above Dundee Dam.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-4. *Chironomus dilutus* survival and biomass results compared with urban reference data

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL



Hyalella azteca survival
- Outside reference envelope
- Within reference envelope

Hyalella azteca biomass
- Outside reference envelope
- Within reference envelope

- Remediated location[a]
- Dredge zone
- Bridge
- Abutment
- Dock

- River mile
- Mudflat
- Gravel with fines
- LPRSA

0 0.25 0.5 Miles

0 0.25 0.5 Kilometers

[a]Two samples were collected in the RM 10.9 dredge area.

Note: Data are control-normalized; reference data are from the area of the Passaic River above Dundee Dam.

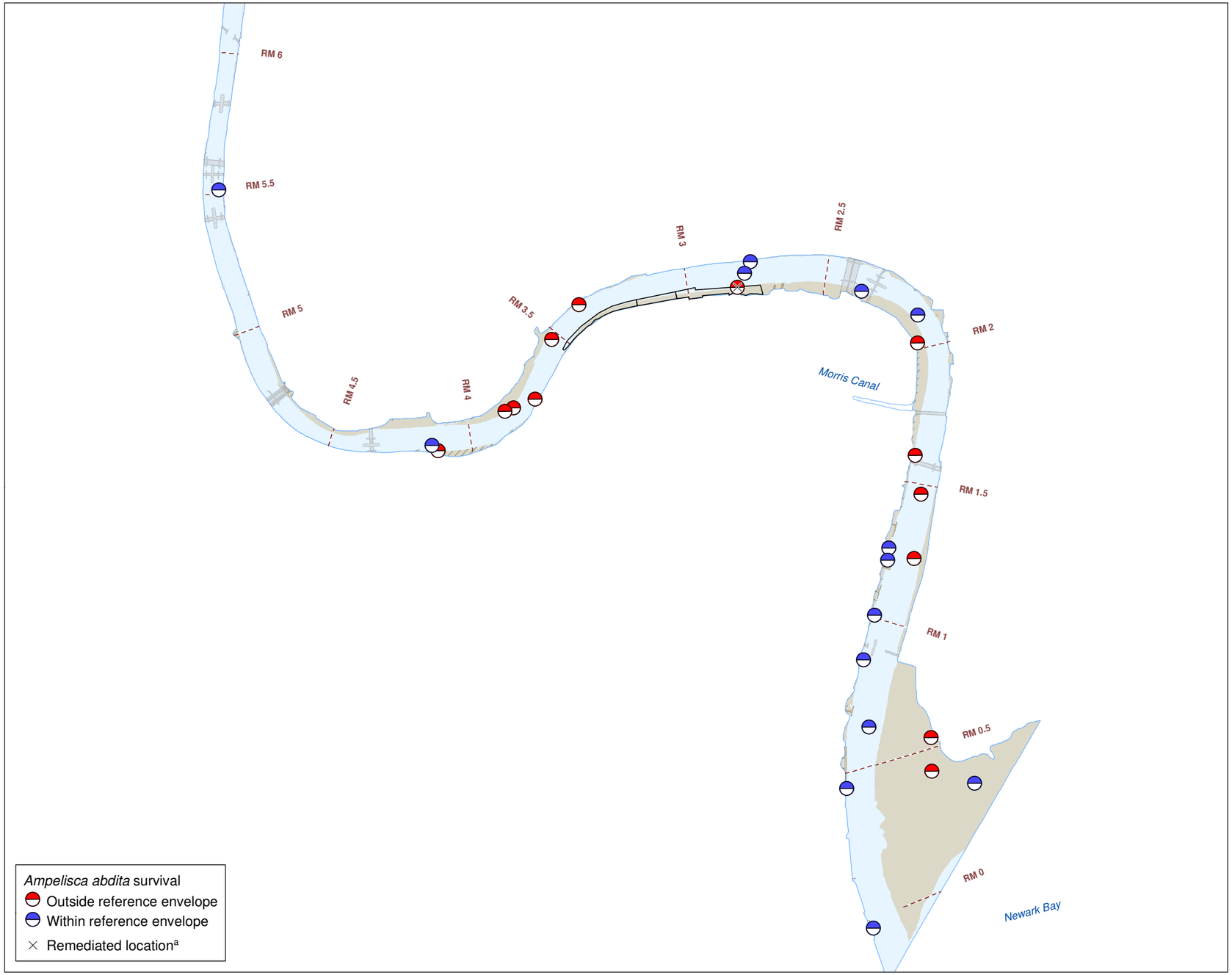
Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-5. *Hyalella azteca* survival and biomass results from freshwater LPRSA locations compared with urban reference data

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

User Name: mikey 5/30/2019, W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6375_Freshwater Hyalella results_LSM_20160408.mxd



[a]One sample was collected in the Lister Ave. dredge area at RM 2.8

Note: Data are control-normalized; reference data are from Mullica River and Great Bay.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-6. *Ampelisca abdita* survival results from estuarine LPRSA locations compared with non-urban reference data
Appendix P. Lower Passaic River Study Area
Baseline Ecological Risk Assessment
FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6443_Ampelisca non-urban estuarine results_LSM_20160721.mxd

Table 3-6. Summary of comparison of sediment toxicity test results to reference data

Parameter	Toxicity Test Endpoints Less than Reference Envelope Threshold (5th Percentile Value)							
	Estuarine[a]				Freshwater[a]			
	A. abdita		*H. azteca*		*C. dilutus*		*H. azteca*	
	Survival		Survival	Biomass	Survival	Biomass	Survival	Biomass
Total no. of locations in LPRSA	27		27	27	71	71	71	71
Reference dataset	Jamaica Bay (urban)	Mullica River and Great Bay (non-urban)	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]
Reference envelope threshold (% of control)	84.5	85.2	75.8	36.2	82.7	77.8	75.8	36.2
Range of LPRSA results (% of control)	15–97		7–99	3–96	16–106	5–104	16–105	6–115
No. of locations below reference envelope threshold	12 (44%)	13 (48%)	11 (41%)	7 (26%)	13 (18%)	58 (82%)	39 (55%)	27 (38%)

[a] "Estuarine" and "freshwater" in Table 3-6 refer to types of sediment toxicity tests defined by interstitial salinity at the time of sediment sampling for toxicity tests. Estuarine is defined as salinity < 5 ppt, and freshwater is defined as salinity ≥ 5 ppt.

[b] *H. azteca* sediment toxicity data were not available from Jamaica Bay; estuarine toxicity data from *H. azteca* (salinity-acclimated) tests were compared with reference data from the area above Dundee Dam.

LPRSA – Lower Passaic River Study Area

ppt – parts per thousand

3.2.4 Summary of results

Based on the comparison of LPRSA data to reference area data, it appears that there is sediment toxicity at many LPRSA locations; however, sediment toxicity does not always correspond between endpoints (e.g., toxicity for both species) at the same location (Figures 3-3 to 3-6; Appendix B, Tables B3-6, B3-8, and B4-5). LPRSA toxicity exceeding the reference envelope was prevalent for *C. dilutus* biomass (82% of freshwater locations) and *H. azteca* and *A. abdita* survival; *H. azteca* survival exceeded the urban reference conditions at 41 and 55% of estuarine and freshwater LPRSA locations, respectively, and *A. abdita* survival exceeded the urban and non-urban estuarine reference conditions at 44 and 48% of estuarine LPRSA locations, respectively. However, neither *C. dilutus* biomass nor *A. abita* survival in the LPRSA were significantly less than their respective reference area condition (Table 3-5). Similarly, *H. azteca* survival was not significantly less in estuarine LPRSA locations than in the area above Dundee Dam, although *H. azteca* survival in freshwater LPRSA locations was significantly less than survival in the area above Dundee Dam.

The *C. dilutus* survival endpoint was the least sensitive among those tested, with only 18% of freshwater LPRSA locations outside of the reference envelope. Biomass of *H. azteca* (for both endpoints and both toxicity test location types) was moderate; 26 and 38% of LPRSA estuarine and freshwater locations were outside the reference envelope, respectively. The *H. azteca* biomass endpoint was low in acceptable urban freshwater reference locations (36 to 47% of negative control), suggesting that this endpoint was influenced by urban stressors. Based on these rates, it appears that LPRSA sediments are often toxic relative to reference conditions, but that uncertainty still exists for this LOE due to inconsistent results across toxicity test endpoints.

3.2.5 Uncertainties in comparison to reference data

A variety of uncertainties associated with the benthic sediment toxicity tests could affect the evaluation of test data and the interpretation of comparisons to reference data.

- There are inherent uncertainties associated with using a reference condition approach to characterize risk. Namely, reference areas are not exactly similar to the LPRSA in terms of biological community or physical and chemical conditions. Reference conditions are used as a model for relatively uncontaminated conditions, but models are imperfect. Regardless, reference conditions provide a reasonable baseline for comparison, assuming that the reference areas are similar enough to the LPRSA. The data used are described in more detail in Section 4.2 of the main BERA text; also, Appendix L, Attachment L3, provides figures that show similarities and differences between LPRSA and reference area habitat-variable and chemical concentration datasets.

- It is unclear whether the screened reference area datasets accurately reflect the reference condition. Specifically, screening reference area data using sediment chemistry and sediment toxicity criteria impose a potentially unreasonable constraint on data acceptability. The resulting datasets may not capture the full range of possible sediment toxicity test results that should be expected under urbanized conditions (but for the LPRSA-specific release of hazardous materials).
- It is unclear whether the comparison of LPRSA data to data from a non-urban reference area (Mullica River and Great Bay) is relevant for characterizing risks in the LPRSA. Comparison to non-urban conditions fails to incorporate potential stressors that are generally observed in urban settings and are expected to influence the LPRSA benthic invertebrate community. Examples of these stressors include altered hydrology due to channelization and flood controls and increased organic and inorganic inputs from CSOs, SWOs, road waste, and permitted industrial discharges.
- The screening of acceptable freshwater reference locations is presented in Appendix B, Tables B4-2 and B4-3. Only five samples from above Dundee Dam were deemed acceptable, adding substantial uncertainty to the quantification of a reference condition.
- Inconsistencies in the observed sensitivities of invertebrate species may be due, in part, to different sensitivities to specific chemicals (Phipps et al. 1995). However, some research has shown that *C. dilutus* and *H. azteca* have similar sensitivities (Ingersoll et al. 2015).
- In order to minimize the possibility of changing the characteristics of the sediment, the sediment was not sieved in the field or by the laboratory staff prior to toxicity testing. As a result, indigenous organisms present in the samples that were too small to be observed without a microscope may not have been removed prior to testing. Thus, the survival and biomass endpoint results could have been influenced by the presence of potential predators and/or non-test organism amphipods or chironomids. If such organisms had been visible during toxicity testing, it is assumed that the toxicology laboratory would have noted the presence of non-test organisms.
- The grain size of the sediment samples collected from the LPRSA varied from fine to coarse, and typically benthic organisms have been shown to have a preference for a particular sediment particle size (USEPA and USACE 1998; Relyea et al. 2012). However, both *H. azteca* and *C. dilutus* tolerate a wide range of sediment grain sizes and types of organic matter (ASTM 2010; USEPA 2000), and are appropriate test organisms for use when the grain sizes of the sediment sample vary widely. To evaluate the effect of grain size on the survival of *C. dilutus* and *H. azteca*, concurrent toxicity tests were conducted using a range

of grain sizes (i.e., 50, 60, and 70% artificially prepared coarse substrate). The results of the tests indicated that the survival of both *C. dilutus* and *H. azteca* in the coarser substrates was similar to that of the negative control.[21]

- Because the sediment samples were collected over an eight-week time period, the toxicity testing was conducted in batches to avoid prolonging the time between sample collection and testing. Differences in test organisms and slight differences in test conditions could have introduced some uncertainty into interpretation of the results. Control-normalized toxicity values were reported in this assessment in order to minimize potential uncertainty caused by batch variability.
- *A. abdita* cannot be cultured in the laboratory and must be field collected (Windward 2018b). Thus, the health of the population is dependent on conditions at the collection site, and consistent results are not guaranteed. The uncertainty associated with using field-collected organisms was reduced by following strict QA/QC procedures, such as the requirement to meet the 90% negative control survival test acceptability criterion.
- An uncertainty associated with the *C. dilutus* biomass endpoint stems from the large number of larvae that pupated and emerged during the tests, indicating that the specimens might have been older than 2nd to 3rd instar at the start of testing. To prevent pupation, Mount (2011) advocated for a maximum starting weight criterion of 0.12 mg ash-free dry weight (AFDW) per organism. This weight is less than the starting weights in tests with LPRSA and above Dundee Dam sediments, which ranged from 0.062 to 0.401 mg/organism, with an average of 0.25 mg/organism across batches and study areas. Unfortunately, this recommendation was not available until after the standard testing protocols had been approved for use, and after all toxicity tests for LPRSA sediment had been completed. The standard protocols used to test toxicity in sediments from the LPRSA and above Dundee Dam were approved by USEPA, and were consistent with current guidance at that time. Moreover, the test organism supplier (Aquatic Bio Systems Inc., Fort Collins, Colorado) and test facility (EnviroSystems Inc., Hampton, New Hampshire) verified that the larvae were either 2nd or 3rd instar, and that they were not expected to emerge until after the 10-day exposure.

  The number of pupated and emerged individuals was counted and recorded on laboratory bench sheets, which were attached to the laboratory reports appended to both the fall 2009 sediment toxicity data report for the LPRSA (Windward 2018b) and the 2012 sediment toxicity reference data report for above Dundee Dam (Windward 2018a). Appendix K, Table K5, provides the

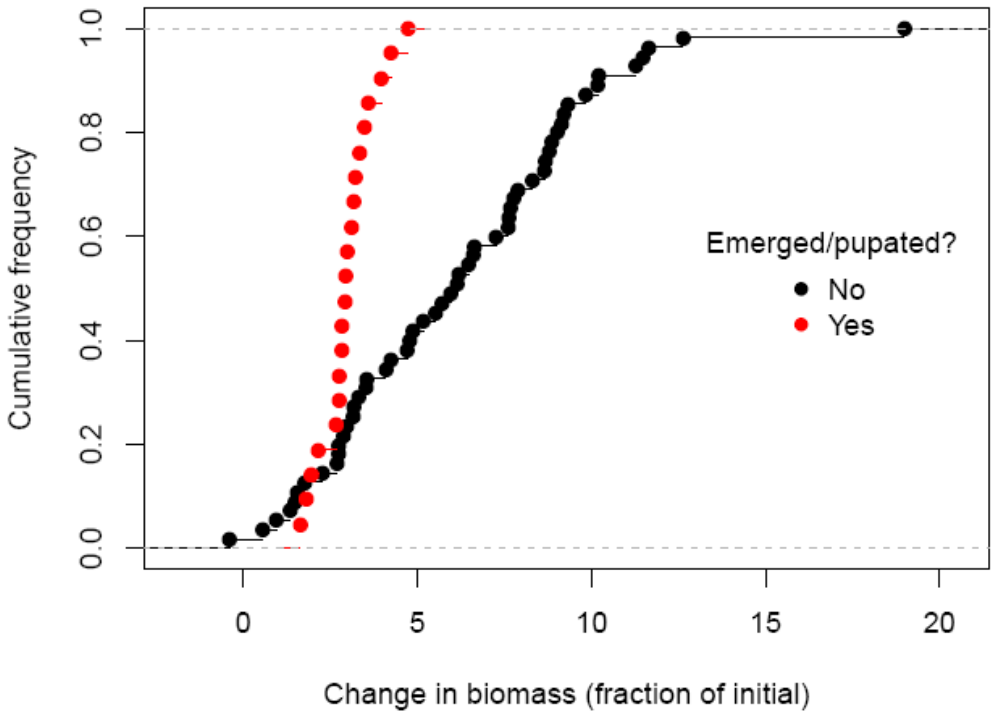
[21]A test evaluating the grain size for *A. abdita* was not conducted as a specific, separate test during this investigation.

combined number of pupated and emerged individuals for each replicate of the LPRSA tests, and Appendix L, Attachment L1, Table L8 provides the same type of data for the area above Dundee Dam. *C. dilutus* biomass was calculated by dividing the ash-free dry weight of surviving larvae for each replicate by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.[22]

C. dilutus toxicity testing using LPRSA sediment samples was conducted in five batches. Organisms emerged or pupated in two of the five batches. In Batch 5, pupation/emergence occurred in 55 of 120 replicates for 13 of the 15 samples, including the laboratory control. In Batch 6, pupation/emergence occurred in 23 of 112 replicates for 8 of the 14 samples, including the laboratory control. In tests for above Dundee Dam sediments, pupation/emergence occurred in 111 of 200 replicates for 24 of the 25 samples, including the laboratory control. As many as seven organisms emerged in each replicate (across all samples and the two study areas).

Figure 3-7 shows the distribution of mean biomass change (as a percent of the starting mass) in LPRSA sediment toxicity tests. On average, growth was significantly lower on samples in which there was emergence/pupation (in any replicate).

[22] It would have been inappropriate to include the weight of *C. dilutus* pupae and adults in the biomass calculations because mass was lost during pupation. It would also have been inappropriate to divide the weight of surviving *C. dilutus* larvae by the initial number of organisms without excluding the number that pupated or emerged; that would have decreased the calculated biomass as if the pupated/emerged organisms had died.



Note: change in biomass is reported as (final biomass – initial biomass)/initial biomass

Figure 3-7. Change in *C. dilutus* biomass in tests with LPRSA sediments and negative controls

- The use of *H. azteca* with sediment representative of estuarine locations may have introduced uncertainty into the interpretation of the results. Although *H. azteca* is tolerant of brackish conditions (USEPA 2000) and can be tested using sediment with a salinity of up to 15 ppt (ASTM 2010), organisms may be under more stress at a salinity of 10 ppt (even after a period of acclimation to a higher salinity) than those tested in freshwater.
- The use of a static-renewal exposure condition when testing *A. abdita* may have introduced a slight positive bias to results (relative to a static exposure) due to volatilization of certain hazardous substances. This is expected to be a minor point of uncertainty because volatile chemicals were often below detectable concentrations in LPRSA sediments (Appendix K).
- Comparisons of LPRSA toxicity test results to negative control results did not account for matrix effects (e.g., the influence of sediment grain size, TOC, or other sediment characteristics); it cannot be stated with certainty whether significant differences in toxicity between LPRSA sediments and negative control sediments were caused by hazardous substances, by other non-chemical factors, or by a combination of chemical and non-chemical factors.
- Interlaboratory variability may have caused discrepancies between, for example, the sensitivities of test species used to develop different datasets.

Reasons for these discrepancies may be related to the use of different laboratories with different technicians, exposure conditions, and laboratory cultures. This uncertainty was most pronounced for the comparison of *A. abdita* data from estuarine LPRSA toxicity locations (i.e., samples with ≥ 0.5 ppt interstitial salinity) to data from Jamaica Bay or Mullica River and Great Bay. Similar comparisons of LPRSA *H. azteca* and *C. dilutus* data to data from above Dundee Dam were less uncertain because a single laboratory conducted all the tests. Control normalization of test results partly addressed this uncertainty by standardizing toxicity test results to the health of laboratory cultures (reflected by negative control result) at the time of sampling. Similarly, test acceptability criteria were met by both laboratories, indicating that each test culture was healthy. Test acceptability criteria are assumed to have been met by the laboratory(ies) used to test reference area sediment toxicity on behalf of USEPA REMAP (1998, 2002a, 1993).

- Source cultures for *H. azteca* and *C. dilutus* tested in LPRSA sediments (fall 2009) likely differed from those tested for *H. azteca* and *C. dilutus* in sediments from above Dundee Dam (fall 2012). Control-normalization of test results partly addressed this uncertainty. Similarly, test acceptability criteria were met during both testing events, indicating that each test culture was healthy.
- Mullica River freshwater data was not used to characterize risk, because the only toxicity test used on sediments in that area tested *A. abdita*, and these tests were not directly comparable to freshwater LPRSA tests with *H. azteca* or *C. dilutus*. Survival of *A. abdita* in Mullica River freshwater samples (after screening data at a mPECq of 0.5 and survival result of 75% of negative control)[23] ranged from 85.1 to 89.5% of negative control survival; survival of *H. azteca* (another amphipod) above Dundee Dam ranged from 76 to 94% of control, and survival of *H. azteca* in the LPRSA ranged from 45 to 100%. These ranges were similar.

3.2.5.1 Quantitative analysis of uncertainty

A quantitative analysis was conducted to address several key uncertainties associated with the sediment toxicity LOE. The methods and results are presented below.

Methods

The 1–90th percentile minimum significant difference (MSD) threshold method, developed by Phillips et al. (2001), was used to evaluate LPRSA sediment toxicity data relative to negative control results for the quantitative analysis of uncertainty. The 1–90th percentile MSD is a site-specific threshold that corresponds to a single toxicity test value below which there is a 90% probability of discerning a statistically significant

[23] One sample from the Mullica River freshwater dataset was removed during the toxicity test screen; *A. abdita* survival in that sample was 41.8% of negative control.

difference between a LPRSA toxicity test value and the negative control. The 1–90th percentile MSD was calculated according to the approach outlined by Phillips et al. (2001), modified to use the MSD calculation presented by Zar (1996). The analysis is presented in Appendix B, Table B5-2. Specifically, a MSD value was calculated for every LPRSA sample (using the raw replicate test data) and the batch-specific negative control data. The MSD for each sample was then normalized to the batch-specific negative control, and the 90th percentile of all control-normalized MSDs was calculated. The 1-90th percentile MSD threshold was simply 1 minus the 90th percentile MSD. Control-normalized toxicity test results from LPRSA locations could thus be compared to the calculated threshold.

For quantitative comparison to reference area data, extreme reference area sediment toxicity test results were removed rather than imposing toxicity screening criteria (Appendix B, Tables B3-3, B4-3). This approach was intended to capture the natural variability in the reference area SQT datasets while excluding data that are clearly different than the rest of the reference area data. This is discussed in more detail in Section 2.3.1 and shown in Figure 2-1.

Results

For the Jamaica Bay dataset, *A. abdita* survival in the 45 samples that passed both the initial chemical screening step and the toxicity screen of extreme values (Appendix B, Table B3-3) ranged from 38 to 108% of the negative control results, with a mean control-normalized survival of 87.5% (Table 3-7).

Table 3-7. Summary of *A. abdita* survival data (fraction of control) from reference areas for the quantitative analysis of uncertainty

Statistic	Jamaica Bay	Mullica River and Great Bay
Sample size	45	12
Minimum	0.380	0.844
Maximum	1.08	1.07
Mean	0.875	0.959
Standard deviation	0.158	0.069
5th percentile[a]	0.558	0.852
10th percentile	0.688	0.861
25th percentile	0.811	0.917
Median	0.930	0.984
75th percentile	0.980	1.00
90th percentile	1.01	1.01
95th percentile	1.03	1.04

Source: USEPA (2016); NOAA (2013)

Note: Three samples were removed from the Jamaica Bay dataset based on best professional judgment regarding low *A. abdita* survival values (i.e., JB310, JB315, and JB366). Survival in sediments from those locations ranged from 1.1 to 11% of the negative control. No values were identified as extreme using the three-times-IQR approach described in Section 2.3.4.1.

[a] Reference envelope threshold for A. abdita survival.

IQR – interquartile range
NOAA – National Oceanic and Atmospheric Administration
USEPA – US Environmental Protection agency

The dataset from above Dundee Dam for the quantitative analysis of uncertainty was substantially larger than that used to analyze the sediment toxicity LOE; the dataset for the uncertainty analysis consisted of 15 samples (an increase from 5 samples). No extreme low toxicity values were identified or removed based on sediment toxicity (Appendix B, Table B4-3). Table 3-8 presents the summary of sediment toxicity data from above Dundee Dam for the quantitative analysis of uncertainty.

Table 3-8. Summary of toxicity reference data (fraction of control) from the area above Dundee Dam for the quantitative analysis of uncertainty

Statistic	*C. dilutus*		*H. azteca*	
	Survival	Biomass[a]	Survival	Biomass[b]
Sample size	15	15	15	15
Minimum	0.71	0.637	0.49	0.261
Maximum	0.96	1.03	0.94	0.518
Mean	0.83	0.758	0.75	0.398
Standard deviation	0.068	0.0894	0.11	0.084
5th percentile[c]	0.72	0.668	0.60	0.263
10th percentile	0.73	0.686	0.65	0.272
25th percentile	0.80	0.712	0.68	0.335
Median	0.83	0.742	0.73	0.414
75th percentile	0.87	0.783	0.82	0.463
90th percentile	0.90	0.816	0.88	0.482
95th percentile	0.92	0.881	0.91	0.497

Source: Windward (2018a)

Note: no extreme values were observed or removed from the dataset from above Dundee Dam.

[a] Biomass for *C. dilutus* was calculated as the total AFDW for each replicate divided by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.

[b] Biomass for *H. azteca* was calculated as the total weight for each replicate divided by the initial number of organisms introduced into the test chamber.

[c] Statistic was the reference envelope threshold value for *C. dilutus* and *H. azteca* toxicity test endpoints used for the quantative analysis of uncertainty (of the sediment chemistry LOE).

AFDW – ash-free dry weight
LOE – line of evidence
Windward – Windward Environmental LLC

MSD values were calculated for each LPRSA sediment sample and toxicity test endpoint (Appendix B, Table B5-2). The 1–90th percentile MSD threshold for each

endpoint is provided in Table 3-9. Thresholds range from 70 to 81% of the negative control data. Table 3-10 summarizes the results of the 1–90th percentile MSD analysis and Figure 3-8 shows the distribution of results in the LPRSA. The results suggest that LPRSA sediments relative to the negative control are less toxic than suggested by comparisons of those data using t-tests (Section 3.2.1).

Table 3-9. 1–90th percentile MSD values for LPRSA toxicity tests

Toxicity Test Endpoint	1–90th Percentile MSD (% of Control)[a]
C. dilutus biomass	71
H. azteca biomass	73
C. dilutus survival	74
H. azteca survival	70
A. abdita survival	81

[a] 1–90th percentile MSD thresholds are based on control-normalized MSDs; this is to facilitate comparison of LPRSA data (also control normalized) to MSD-based thresholds.

LPRSA – Lower Passaic River Study Area

MSD – minimum significant difference

Table 3-10. Summary of comparison of LPRSA toxicity test results to 1–90th percentile MSD thresholds

Toxicity Test Endpoint	No. of LPRSA Locations	No. of LPRSA Locations Below 1–90th Percentile MSD
C. dilutus biomass	71	50 (70%)
C. dilutus survival	71	5 (7%)
H. azteca biomass	98	88 (90%)
H.azteca survival	98	43 (44%)
A. abdita survival	27	8 (30%)

LPRSA – Lower Passaic River Study Area

MSD – minimum significant difference

RM 2.5
RM 2
Morris Canal
RM 1.5
RM 1
RM 0.5
RM 0
Newark Bay
N
1
RM 5
RM 4.5
RM 4
RM 3.5
RM 3
RM 2.5
Morris Canal
N
2
RM 6.5
RM 6
RM 5.5
RM 5
N
3
RM 9
RM 8.5
Second River
RM 8
N
4
RM 13
RM 12.5
RM 12
Third River
RM 11.5
RM 11
RM 10.5
N
5
Dundee Dam
RM 17
RM 16.5
RM 16
RM 15.5
RM 14.5
Saddle River
RM 14
RM 15
RM 13.5
N
6
6
5
4
3
2
1
N
Map frame positions

Toxicity test type
Estuarine
Hyalella Biomass
Hyalella Survival
Ampelisca Biomass (no data)
Ampelisca Survival
Freshwater
Hyalella Biomass
Hyalella Survival
Chironomus Biomass
Chironomus Survival

MSD result
< 1-90th percentile
≥ 1-90th percentile

× Remediated location[a]
Dredge zone

River mile
Bridge
Abutment
Dock
Mudflat
Gravel with fines
LPRSA

0 0.25 0.5 Miles
0 0.25 0.5 0.75 Kilometers

[a]One sample was collected in the Lister Ave. dredge area at RM 2.8 and two were collected in the RM 10.9 dredge area.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Windward environmental LLC

Figure 3-8. Results from comparisons of LPRSA toxicity test data to negative control results for the quantitative analysis of uncertainty

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

Prepared by mikey 5/30/2019, W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6451_Bioassay negative control results-alternative-_LSM_20160721.mxd

The results of the Mann-Whitney U tests comparing LPRSA sediment toxicity data to reference area data (for the quantitative analysis of uncertainty) are presented in Table 3-11 and Appendix B (Tables B6-2c and B6-2d). The results are similar to those of the approach used for the sediment toxicity LOE (Table 3-5; Appendix B, Tables B6-2a and B6-2b), with the exception that *A. abdita* survival is significantly different in the LPRSA than in both Jamaica Bay (urban) and Mullica River/Great Bay (non-urban). The lack of a significant result associated with the LOE approach (Table 3-5) is due not to substantially lower survival, but rather to the small sample size of the reference area dataset developed using the LOE approach (n = 5); median *A. abdita* survival values from each dataset (Tables 3-4 and 3-7) are very similar (i.e., 93 to 98% survival).

Table 3-11. Results of the statistical analysis comparing the LPRSA toxicity test data with reference data for the quantitative analysis of uncertainty

Area of Interest	Toxicity Test Type	Is there a significant difference between LPRSA and reference area data? (Mann-Whitney U Test, alpha = 0.05)[a]				
		C. dilutus **Survival (% of Control)**	***C. dilutus*** **Biomass (% of Control)[b]**	***H. azteca*** **Survival (% of Control)**	***H. azteca*** **Biomass (% of Control)[c]**	***A. abdita*** **Survival (% of Control)**
Jamaica Bay	estuarine	nt	nt	nt	nt	**yes; p = 1.01E-02**
Mullica River and Great Bay	estuarine	nt	nt	nt	nt	**yes; p = 1.73E-04**
Above Dundee Dam	estuarine[d]	nt	nt	no; p = 7.27E-01	no; p = 9.89E-01	nt
Above Dundee Dam	freshwater	no; p = 8.77E-01	no; p = 2.01E-01	**yes; p = 2.87E-02**	no; p = 9.07E-01	nt

Note: Mann-Whitney U test conducted as a one-tailed test, $p < 0.05$. An "nt" value is reported for tests that were not conducted for the given test species in the given reference area and/or in the LPRSA.

Bold text indicates significant result ($p < 0.05$).

[a] Based on the Mann-Whitney U Test.

[b] Biomass for *H. azteca* was calculated as the total weight for each replicate divided by the initial number of organisms introduced into the test chamber.

[c] Biomass for *C. dilutus* was calculated as the total AFDW for each replicate divided by the initial number of organisms introduced into the test chamber minus the number of organisms that either emerged or pupated during the test.

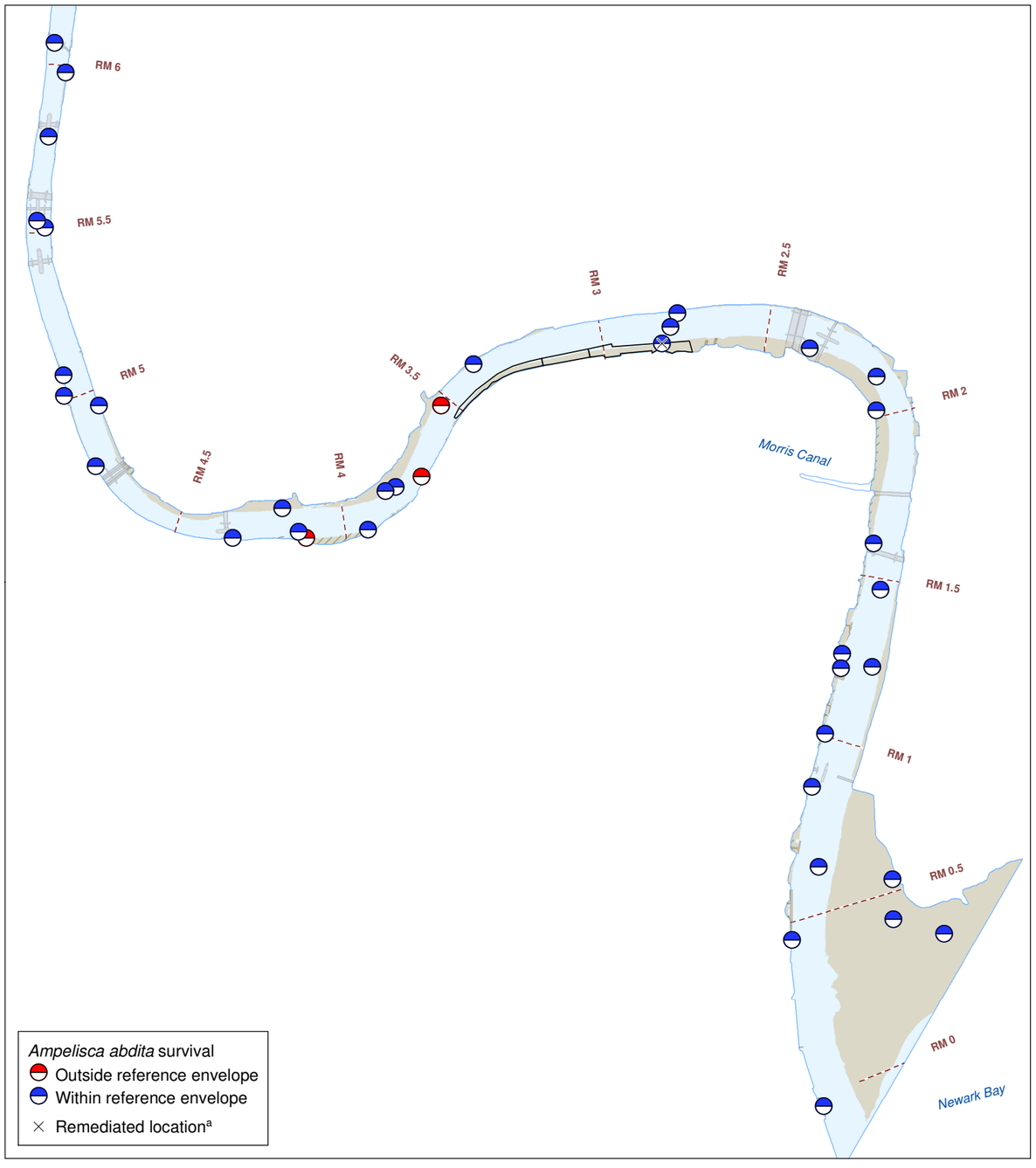
[d] *H. azteca* test results from the estuarine LPRSA could not be compared to estuarine reference area data; instead, they were compared to freshwater *H. azteca* test results from above Dundee Dam.

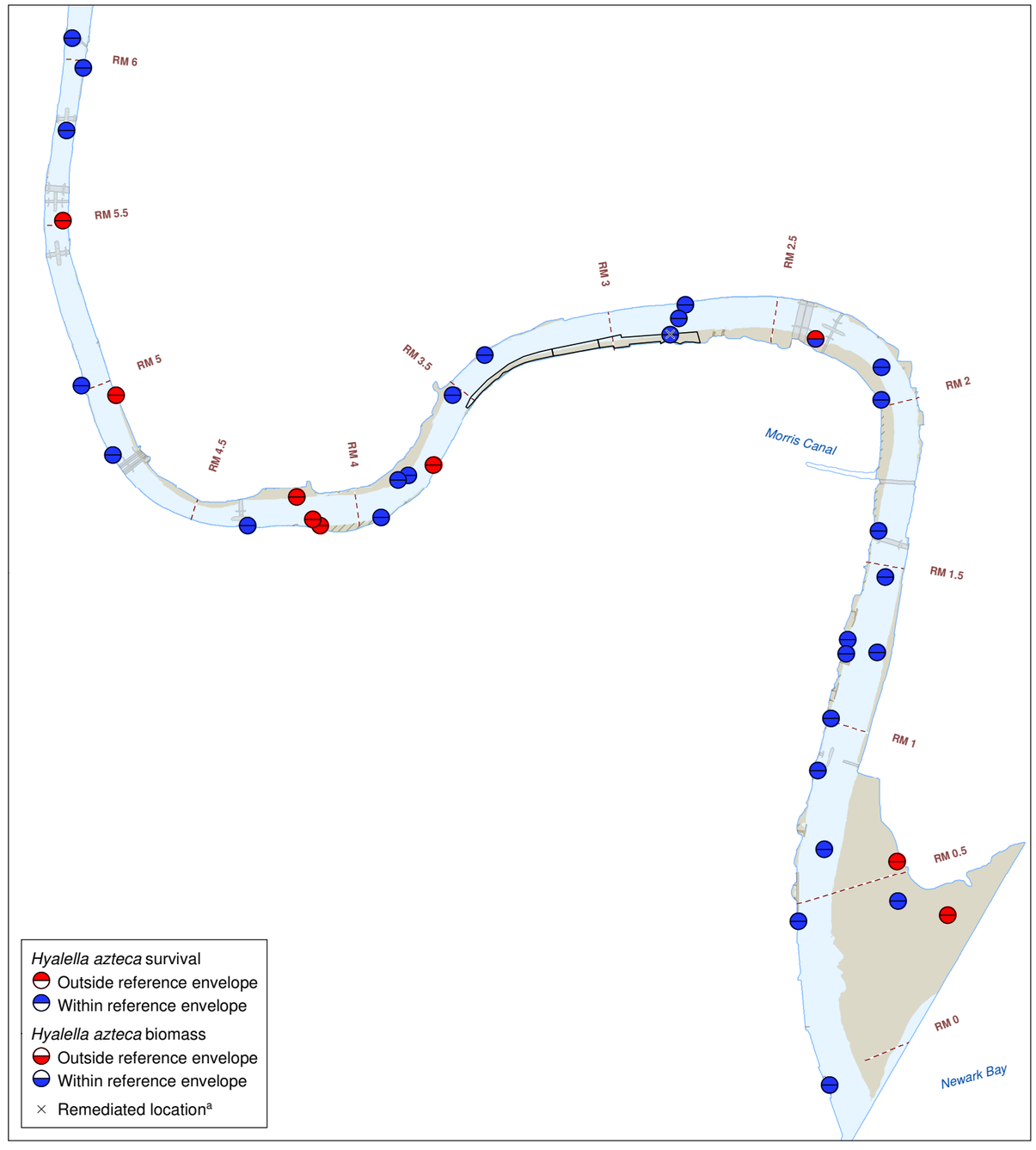
AFDW – ash-free dry weight LPRSA – Lower Passaic River Study Area

nt – not tested

The location-by-location comparisons between sediment toxicity data from LPRSA SQT locations and reference envelope thresholds (for the quantitative analysis of uncertainty) are shown in Figures 3-9 through 3-11 and summarized in Table 3-12. The full analysis is presented in Appendix B, Tables B3-7, B3-9, and B4-6. The quantitative analysis of uncertainty resulted in substantially lower rates of reference envelope exceedance for toxicity test data (Table 3-12). For example, *A. abdita* survival exceeded envelope conditions at only 11% of estuarine LPRSA locations (Table 3-12), compared

to 44% for the LOE approach (Table 3-6). Although *A. abdita* survival exceeded the reference threshold at fewer LPRSA locations based on the quantitative analysis of uncertainty, *A. abdita* survival in the LPRSA was significantly lower than the reference condition (Table 3-11).





Windward environmental LLC

N

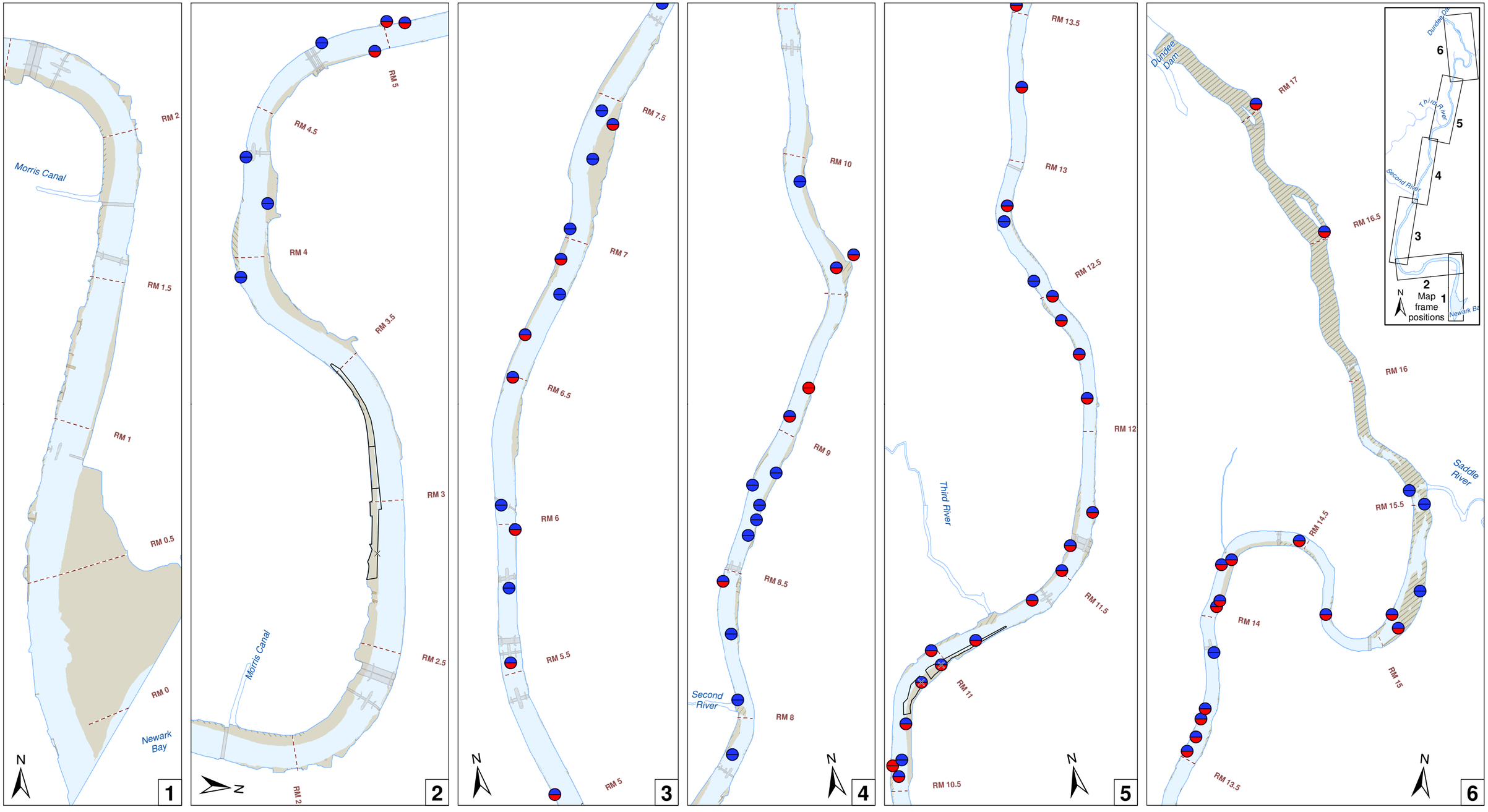
0 0.25 0.5 Mile

Dredge zone
Bridge
Abutment
Dock
River mile
Mudflat
Gravel with fines
LPRSA

[a]One sample was collected in the Lister Ave. dredge area at RM 2.8
Note: LPRSA *A. abdita* toxicity data were compared to urban reference data from Jamaica Bay; LPRSA *H. azteca* toxicity data were compared to urban reference data from the area of the Passaic River above Dundee Dam.
Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-9. *Ampelisca abdita* survival and *Hyalella azteca* survival and biomass results from estuarine LPRSA locations compared with urban reference data for the quantitative analysis of uncertainty
Appendix P. Lower Passaic River Study Area
Baseline Ecological Risk Assessment
FINAL

Prepared by mikey 5/30/2019; W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6372_Ampelisca Hyalella urban estuarine results -alternative-_LSM_20160721.mxd



Windward environmental LLC

0 0.25 0.5 Miles

0 0.25 0.5 Kilometers

Chironomus dilutus survival
- Outside reference envelope
- Within reference envelope

Chironomus dilutus biomass
- Outside reference envelope
- Within reference envelope

- Remediated location[a]
- Dredge zone
- Bridge
- Abutment
- Dock

- River mile
- Mudflat
- Gravel with fines
- LPRSA

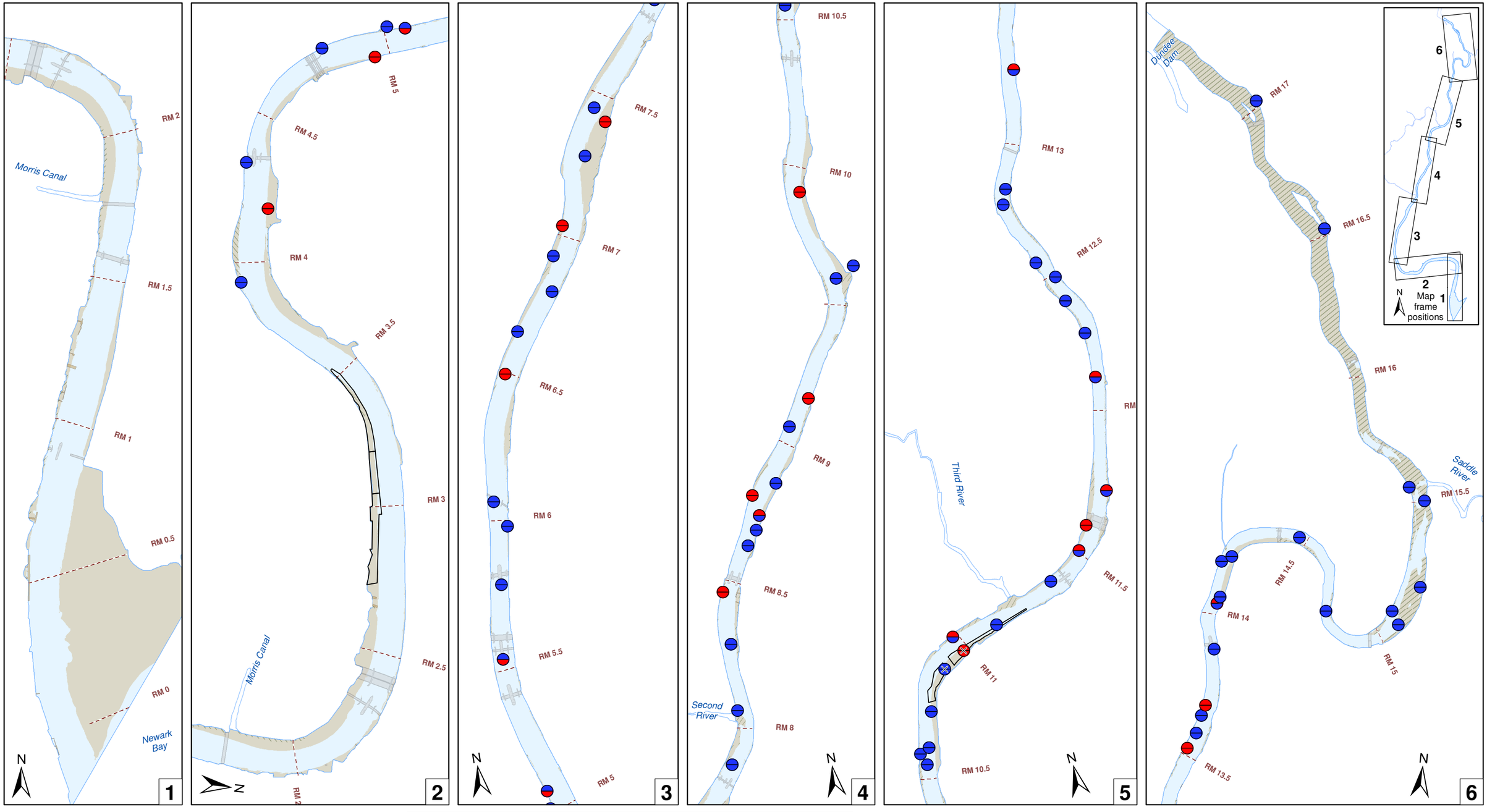
[a]Two samples were collected in the RM 10.9 dredge area.

Note: Data are control-normalized; reference data are from the area of the Passaic River above Dundee Dam.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Figure 3-10. *Chironomus dilutus* survival and biomass results compared with urban reference data for the quantitative analysis of uncertainty
Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment
FINAL

User Name: mikey 5/30/2019, W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6374_Chironomus FW results -alternative-_LSM_20160603.mxd



Hyalella azteca survival
- Outside reference envelope
- Within reference envelope

Hyalella azteca biomass
- Outside reference envelope
- Within reference envelope

× Remediated location[a]

Dredge zone

Bridge

Abutment

Dock

River mile

Mudflat

Gravel with fines

LPRSA

[a]Two samples were collected in the RM 10.9 dredge area.

Note: Data are control-normalized; reference data are from the area of the Passaic River above Dundee Dam.

Mudflat areas for ecological receptors are those areas where the river bottom slope is ≤ 6° and the depth is ≥ -2 ft MLLW, based on the 2007 CPG bathymetric survey.

Windward environmental LLC

0 0.25 0.5 Miles

0 0.25 0.5 Kilometers

Figure 3-11. *Hyalella azteca* survival and biomass results from freshwater LPRSA locations compared with urban reference data for the quantitative analysis of uncertainty

Appendix P. Lower Passaic River Study Area Baseline Ecological Risk Assessment

FINAL

User Name: mikey 5/30/2019, W:\Projects\06-58-01 Passaic RI\Data\GIS\Maps_and_Analysis\BERA\Revised BERA 2016\6376_Freshwater Hyalella results-alternative-_LSM_20160603.mxd

Table 3-12. Summary of comparison of sediment toxicity test results to reference data for the quantitative analysis of uncertainty

Parameter	Toxicity Test Endpoints Less than Reference Envelope Threshold (5th Percentile Value)							
	Estuarine[a]				Freshwater[a]			
	A. abdita		*H. azteca*		*C. dilutus*		*H. azteca*	
	Survival		Survival	Biomass	Survival	Biomass	Survival	Biomass
No. of LPRSA samples	27		27	27	71	71	71	71
Reference dataset	Jamaica Bay (urban)	Mullica River and Great Bay (non-urban)	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]	area above Dundee Dam (urban)[b]
Reference envelope threshold (% of control)	55.8	85.2	59.8	26.3	72.1	66.8	59.8	26.3
Range of LPRSA results (% of control)	15–97		7–99	3–96	16–106	5–104	16–105	6–115
No. of locations below reference envelope	3 (11%)	13 (48%)	7 (26%)	6 (22%)	2 (3%)	45 (63%)	20 (28%)	15 (21%)

[a] Estuarine and freshwater in Table 3-12 refer to types of sediment toxicity tests defined by interstitial salinity at the time of sediment sampling for toxicity tests. Estuarine is defined as salinity < 5 ppt, and freshwater is defined as salinity ≥ 5 ppt.

[b] *H. azteca* sediment toxicity data were not available from estuarine reference areas; estuarine toxicity data from *H. azteca* (salinity-acclimated) tests were compared with reference data from the area above Dundee Dam.

LPRSA – Lower Passaic River Study Area

ppt – parts per thousand

3.3 SUMMARY

Toxicity tests conducted using LPRSA field-collected sediment measured the effects of exposure to the complex mixture of chemicals, non-chemical stressors, and general physical characteristics of the surface sediment at locations within the LPRSA. The survival endpoints for the amphipods *A. abdita* and *H. azteca* and the chironomid *C. dilutus* are relevant for predicting the health of the benthic community in the LPRSA. The survival of amphipods exposed to field-collected sediment in 10- and 28-day toxicity tests (using *A. abdita* and *H. azteca*, respectively) has been correlated with abundance of amphipods, species richness, and other measures of community structure in the field (Schlekat et al. 1994; Swartz et al. 1994; Long et al. 2001). Using sensitive and representative sediment-dwelling species, sediment toxicity testing can be useful in the evaluation of potential harm to benthic invertebrate communities on a location-by-location basis.

Based on comparisons of LPRSA toxicity test results to negative controls, it appears that LPRSA sediments are generally more toxic than negative control tests conducted using clean, artificially formulated laboratory sediments. Based on these results alone, it is not clear to what degree measured toxicity is caused by hazardous substances in sediments or by a combination of chemical and non-chemical stressors. This lack of clarity is why the characterization of risk in the BERA (using a WOE analysis) incorporated the results of reference area comparisons to LPRSA data rather than negative control comparisons. Negative control performance relative to test acceptability criteria suggested that laboratory cultures were in acceptable health.

The comparison of LPRSA and reference area sediment toxicity test data indicated that LPRSA sediments at many locations were toxic, although toxicity was often inconsistent across species and endpoints at a single location. For example, *C. dilutus* survival, typically a less sensitive endpoint, tended to be similar to reference conditions in most LPRSA sediments tested. Also, sediment toxicity at LPRSA locations was generally not significantly lower than at reference area locations. *A. abdita* and *H. azteca* survival were the most sensitive endpoints tested aside from *C. dilutus* biomass, which, as noted in Section 3.3.4, is more uncertain than other toxicity endpoints. *H. azteca* survival (at freshwater locations) was the only statistically significant sediment toxicity test endpoint (i.e., significantly lower survival than the freshwater urban reference conditions above Dundee Dam).

4 Sediment Chemistry Line of Evidence

Sediment chemical concentrations in LPRSA SQT samples were assessed in the screening-level ecological risk assessment (SLERA) (Appendix A) using sediment criteria provided by USEPA Region 2, consistent with USEPA (2001b) guidance, to select chemicals of potential ecological concern (COPECs). However, as agreed with USEPA, an SQT approach (that included the evaluation of a sediment chemistry LOE) was used to evaluate benthic invertebrate risks in the LPRSA (Section 6 of BERA main text). COPECs determined in Appendix A are acknowledged in Section 4.3.4, but the sediment chemistry LOE conducted herein was not used to determine preliminary contaminants of concern (COCs), as is generally done.[24] Rather, an assessment of the sediment chemistry LOE was used to establish weights in the WOE analysis that was used to characterize chemical risks to LPRSA benthic invertebrates (Section 6.1 of the main BERA text).

The following analyses were conducted as part of the sediment chemistry LOE:

- A bivariate correlation analysis between LPRSA sediment chemical concentrations and toxicity test response and benthic community metric data was conducted.
- A multivariate analysis was conducted to compare habitat variables and ordinations of sediment chemistry data to toxicity test response and benthic community metric data. The relative influence of habitat and chemistry on benthic response variables was evaluated.
- LPRSA sediment chemistry data from SQT locations were compared with logistic regression model-based T20 (20% probability of observing toxicity) and T50 (50% probability of observing toxicity) values.
- Detailed uncertainties are discussed in sections describing simultaneously extracted metals (SEM) and acid volatile sulfide (AVS) (USEPA 2005b), total PAHs (equilibrium partitioning [EqP] approach) (USEPA 2003), and the ability of the T20 and T50 criteria or mean-quotient thresholds to accurately predict site-specific sediment toxicity in the LPRSA. Additional points of uncertainty are also noted (including an acknowledgement of COPECs evaluated directly or indirectly as part of this LOE).

As part of the uncertainty analysis, a quantitative analysis of the approach to screening LPRSA sediment chemistry data was conducted. In the quantitative analysis, mean ERM and PEC quotients from LPRSA SQT locations were compared to mean quotient thresholds that were derived using reference area mean ERM or PEC quotients and

[24] Preliminary COCs are determined for surface water and benthic invertebrate tissue LOEs in Sections 6.2 and 6.3 (and reported in Section 6.4) of the BERA main text.

reference area sediment toxicity test data. More detail is provided in the following sections. Sediment chemistry LOE data tables are provided in Appendix B, Tables B1 (bivariate correlation), B2 (multivariate approach), and B7 (sample-by-sample screen of LPRSA data against T20/T50 or mean quotient thresholds).

4.1 METHODS

4.1.1 Bivariate correlation analysis

Bivariate correlation analyses were performed to assess the strength of relationships between dry weight sediment chemistry (concentrations for single chemicals) and benthic response variables (i.e., single toxicity test endpoints or benthic community metrics). The results of the correlation analyses were used to inform conclusions from the WOE analysis outlined in Section 6.1 of the BERA main text. Correlation analyses were conducted using R (R Core Team 2016), and the results are presented in Appendix B, Table B1.

Correlation analyses are used to determine whether two variables are related in a significant way, such that a consistent change in one variable corresponds with a consistent change in another. This is useful when comparing co-located sediment chemistry and benthic response data to assess the assumption that chemical concentrations lead to sediment toxicity and impairment of benthic communities. However, correlation analysis does not prove that a causative relationship exists between two variables.

The nonparametric Spearman rank correlation test was used because the Shapiro-Wilk test of normality ($\alpha = 0.05$)[25] indicated that few of the distributions of chemical concentrations or biological endpoint datasets were normally distributed. This was verified visually using histograms. Spearman rank correlations are used to determine if significant monotonic relationships exist between the ranks of two variables.[26] A monotonic relationship is one in which an increase in one variable consistently corresponds to an increase (direct relationship) or decrease (inverse relationship) in the second variable, the simplest example being a linear relationship. Many non-linear relationships (e.g., sigmoid or logistic dose-response curves) are also monotonic. Spearman rank correlation is a straightforward statistical method that has been used

[25] Tests of normality were conducted using Addinsoft™ XLSTAT software, Version 2012.3.04. Spearman rank correlation analyses were conducted using R programming (2015, Version 3.2.2, `rcorr` function from `Hmisc` package). Pairwise deletion of missing values was used so that as many data as possible were included in each correlation.

[26] This differs from the parametric Pearson product-moment correlation, which assumes a linear correlation between two variables. Dose-response relationships (in controlled studies) are classicly characterized using non-linear models (e.g., log-logistic or sigmoid curves), so the Pearson product-moment correlation method is expected to be of limited use for evaluating chemical-toxicological relationships.

to evaluate sediment chemistry and benthic response datasets collected in several other water bodies (Anderson et al. 2001; Breneman et al. 2000; Canfield et al. 1994).

The strength of the monotonic relationship between ranks of two paired variables is described by the Spearman rank correlation coefficients (r values). Coefficient values range from -1 (perfect negative relationship) to 1 (perfect positive relationship). Greater absolute values of r indicate stronger (negative or positive) monotonic relationships. Values closer to 0 indicate weak relationships.

There is a high likelihood that correlation analysis, when run many times, will result in a significant correlation between two variables that is not truly significant.Multiple Spearman rank correlation analyses were performed, introducing uncertainty in correlation results, specifically the increased probability of making a type II error (false positive). In order to account for this uncertainty, the Spearman rank correlation analysis was run both with and without Bonferroni correction. Bonferroni correction is used to reduce type II errors by dividing the selected significance value (alpha = 0.05) by the number of comparisons being made (i.e., 104 for the site-wide analysis and 72 for the tidal freshwater analysis).[27] Bonferroni correction increases the likelihood of type I errors (false negatives) by setting a stringent threshold for determining statistical significance. By presenting the results of Spearman rank correlations both with and without Bonferroni correction (Section 4.2.1), uncertainties associated with type I/II errors in significant correlation are addressed. Correlation test results that remain significant after Bonferroni correction (of the significance threshold alpha) have high a certainty of true correlatedness.

Prior to analysis, USEPA Region 2 provided CPG with two lists of COPECs to evaluate using correlation analysis: one for LPRSA data from the tidal freshwater zone and one for site-wide LPRSA data (Table 4-1).

[27] The number of comparisons is based on the number of COPECs identified in Table 4-1 multiplied by the number of benthic response variables available in the respective dataset (13 variables site wide and 12 in tidal freshwater).

Table 4-1. Reduced COPEC lists used for Spearman rank correlation analyses

Tidal Freshwater COPEC List	Site-wide LPRSA COPEC List
Total chlordane	Lead
Phenol	Mercury
Total PAHs	Zinc
Bis-(2-ethylhexyl) phthalate	Total LPAHs
Total DDx	Total HPAHs
Total PCB Congeners	Bis-(2-ethylhexyl) phthalate
	Total DDx
	Total PCB Congeners

Note: Table was provided by USEPA Region 2 to CPG (USEPA 2015b).

COPEC – chemical of potential ecological concern
CPG – Cooperating Parties group
DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
HPAH – high-molecular weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
LPRSA – Lower Passaic River Study Area
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT and 4,4′-DDT)
USEPA – US Environmental Protection Agency

4.1.2 Multivariate analysis

Multivariate statistics were applied to LPRSA data in an effort to evaluate potential relationships among sediment chemical concentrations, habitat conditions, and measureable benthic responses.[28] Principal component analysis (PCA), exploratory factor analysis (EFA), and multiple linear regression (MLR) were used in sequence to evaluate relationships among multiple chemical variables, habitat conditions, and benthic response variables (i.e., sediment toxicity test endpoints and infaunal community metrics). The multivariate statistical analyses described herein were conducted in R (R Core Team 2016), and results are presented in Appendix B, Table B2.

4.1.2.1 Multivariate datasets

To account for data constraints and address uncertainty, two approaches were taken to establish sediment chemistry datasets for multivariate analyses. This resulted in two parallel analyses, which are qualitatively compared in Section 4.2.2. Data constraints arose because, if there were missing sediment chemistry data for any variable, multivariate output could not be generated for a sample with missing values; there were many missing sediment chemistry values or variables in the reference area datasets (excepting data from above Dundee Dam). Additionally, log-transformed

[28] Relationships found using multivariate statistics are correlative but not necessarily causative.

data are typically used to conduct PCAs with variables with right-skewed (e.g., lognormal) distributions, but logarithms cannot be calculated for zero. A subset of the Jamaica Bay dataset (i.e., 1998 sampling event) included large numbers of non-detected values that were reported as zero, so these samples were omitted to allow for log-transformation of the sediment chemistry data. Chemistry data were also centered (to the mean) and scaled (to units of standard deviation).

The intent of the first multivariate dataset and associated analyses (hereafter referred to as Method 1) was to incorporate a reasonably large number of COPEC variables into the PCA while maintaining a large amount of reference area samples. This was accomplished by selecting all COPECs for which there were data in all of the LPRSA or reference area SQT datasets. COPECs were then screened out if they had > 5% missing data across all samples, < 50% detection frequency across all samples or < 25% detection frequency within regional reference datasets (i.e., Jamaica Bay, Mullica River freshwater, and Mullica River and Great Bay), and/or 100% missing data in regional reference datasets. Also, to reduce redundancy, only total DDx (sum of all six dichlorodiphenyltrichloroethane [DDT] isomers [2,4′-dichlorodiphenyldichloroethane (DDD), 4,4′-DDD, 2,4′-dichlorodiphenyldichloroethylene (DDE), 4,4′-DDE, 2,4′-DDT and 4,4′-DDT]) was included, rather than other sums of DDT isomers. Similarly, total PAHs were excluded to reduce redundancy with individual PAH analytes. The screening step resulted in a reasonably large COPEC subset (n = 30 analytes) with chemistry data representing all reference area datasets (and the LPRSA).[29] The 30 analytes were primarily composed of individual metals, PAHs, and organochlorine pesticides (i.e., Dieldrin, alpha-Chlordane, total DDx, and hexachlorobenzene). Due to data constraints, total PCBs, phenol, and bis-(2-ethylhexyl) phthalate—all of which were included in USEPA's correlation analysis subset (Table 4-1)—were excluded from the Method 1 dataset.

The intent of the second multivariate dataset and associated analyses (hereafter referred to as Method 2) was to focus on the subset of COPECs identified by USEPA in its statistical guidance for correlation analysis (Table 4-1) (excepting phenol, which was never detected in LPRSA SQT samples, nor measured in regional reference datasets). The result of this approach was a focus of the multivariate analysis on key COPECs of interest, including total PCBs, total PAHs, and bis-(2-ethylhexyl) phthalate, which were not included in the Method 1 dataset. However, data limitations for the

[29] This included (after removing incomplete records) 97 LPRSA samples, 24 samples from above Dundee Dam, 56 samples from Jamaica Bay, 15 samples from Mullica River and Great Bay, and 4 samples from Mullica River freshwater. No effort was made to pre-screen the reference area datasets for acceptability (e.g., using a chemical screening threshold) prior to conducting multivariate statistics (as was done in Sections 2 and 3 to establish reference conditions).

selected COPECs resulted in the removal of most reference area data; only the reference dataset from above Dundee Dam was retained in the Method 2 dataset.[30]

4.1.2.2 PCA and EFA

Principal Component Analysis

PCA is used to reduce many variables to a small subset of independent variables called principal components (PCs), each of which accounts for a certain portion of the overall variance of the original data. Each PC represents a major gradient in the original dataset, such that a group of highly correlated variables is best described by one PC, while other sets of correlated variables (uncorrelated with the first group) are better described by other PCs. Per USEPA guidance, PCA was used in the analysis herein only to estimate the percent of total variance explained by each PC, which was related to an output of PCA called eigenvalues.[31] PCs were generated in such a way that their eigenvalues were ordered from the first to the last PC (accounting for decreasing amounts of variance). The number of PCs needed to cumulatively account for 90 to 95% of the total variance was determined, and that number was used to guide EFA.

Exploratory Factor Analysis

EFA is a multivariate statistical approach that is conceptually similar to PCA in that it can be used to reduce variables to a small subset of independent variables (i.e., factors), but EFA has some subtle conceptual and mathematical differences. Most notably, EFA handles variance by parsing "common variance" and "unique variance," which are distinguished as variance associated with multiple variables (e.g., chemical analytes) and unexplainable variance specific to each variable, respectively. PCA creates PCs that incorporate both types of variance. Conceptually, EFA and PCA differ in that EFA attempts to create hypothetical "latent variables," which are assumed to exist but may not be directly measurable. Examples of latent variables could be heavy metal concentration or PAHs as composites of chemical analyte data, or enrichment composed of TOC, percent fine-grained sediment, and nutrient concentrations. Subsequent analysis (i.e., confirmatory factor analysis) can be used to test whether variables are associated with a hypothetical latent variable, although this did not occur for the analysis described herein.

Several important statistics are generated when conducting EFA: eigenvalues, loading values, and scores. Eigenvalues are described briefly in the previous section, and loading values and scores are described below.

[30] This includes (after removing incomplete records) 97 LPRSA samples and 24 samples from above Dundee Dam.

[31] The percent variance explained by a PC is equal to its squared eigenvalue, divided by the sum of squared eigenvalues of all PCs.

Loading values describe the strength of associations between original variables (e.g., sediment chemistry data) and factors. Highly correlated variables are expected to have high absolute loading values (closer to 1 or -1) for the same factor, whereas uncorrelated variables are likely to have higher loading values on different factors. Loading values are integral to the interpretation of factor scores, because loading values characterize the latent variables that are described by each factor in terms of the original data. Chemicals with high absolute loading values are closely associated with the factors. Chemicals with positive loading values have factor scores that increase with increasing chemical concentrations, and the converse is also true. In order improve the interpretability of EFA results, loading values can be "rotated" using various methods. Rotating factors alters the loadings of chemicals on factors, ideally improving factor interpretation. The orthogonal rotation method "varimax" is used to rotate factors and improve EFA results interpretability.

As noted, factors are independent variables, and factor scores are the individual data that compose those variables. Factor scores can be used in virtually any manner in which other continuous variables are used (e.g., in subsequent statistical analysis), allowing for the integration of EFA results with MLR.

4.1.2.3 MLR

Model Development

Per guidance from USEPA Region 2 (USEPA 2017), MLRs were developed to explain benthic response variables using (in addition to available habitat variables) the same number of EFA factors as PCs that were required to explain ≥ 95% of the total variance in the Method 1 or Method 2 sediment chemistry datasets. MLR models containing all habitat variables and factors are referred to herein as full models.[32] Additionally, more parsimonious limited combined and limited chemistry models were developed; these models included the same number of factors as PCs that were required to explain ≥ 90% of the total variance in the applicable sediment chemistry dataset (applicable to Method 1 or Method 2). The limited chemistry model did not contain habitat variables. Lastly, habitat-only models, with only the two habitat variables (i.e., TOC and total fines),[33] and null (intercept-only) models were developed. These models provided a baseline for evaluating the relationship of sediment chemistry factors to benthic response variables. Thus, six separate MLR models[34] were developed to explain each

[32] Habitat variables, like chemical variables included in the PCA and EFA, were scaled and centered. The variables were not log-transformed because of the limited range (0 to 100%) over which TOC and total fine-grained sediment could be measured.

[33] Additional habitat variables were considered, but very few parameters were consistently measured between or within reference areas and the LPRSA. For example, DO was consistently measured for regional reference area datasets, but it was not measured at LPRSA or above Dundee Dam sampling locations.

[34] The six models included the null, habitat-only, limited chemistry, limited combined, full chemistry, and full models.

benthic invertebrate community metric and sediment toxicity test endpoint, and these models were quantitatively and qualitatively compared. Moreover, this process was conducted for both Method 1 and Method 2.

Checking MLR Assumptions

Two key assumptions of MLR were checked prior to moving forward with subsequent analyses: normality and homoscedasticity of residuals. These assumptions were evaluated visually in R using normal Q-Q plots and residual-fitted value plots, respectively. Q-Q plots with approximately linear residuals (in normal quantile space) were acceptable as normal, and some deviation from normality toward the tails of the distribution was expected. Residual-fitted value plots should have been evenly but randomly distributed around zero, with little discernible trend in residuals or their variance. If the variance of residuals appeared to increase or decrease in relation to the fitted value, then residuals were heteroskedastic. To correct non-normality or heterskedasticity of residuals, benthic response variables were log-transformed prior to fitting MLR models. The Q-Q and residual-fitted value plots were then created for the new models to check if transformation helped to satisfy the modeling assumptions. If not, then the models with untransformed values were used, and uncertainties were stated.

Leverage-residual plots were also inspected for each model to discern the degree to which subsets of data influenced the regression model. For example, single points with a high degree of leverage could significantly skew a model. Leverage was further analyzed using Cook's distance and DFFITS metrics, described below.

Collinearity and Path Analysis

In some cases, the interpretability of MLR could be affected by including many explanatory variables (i.e., variables that are collinear or highly correlated). This was not generally a problem when conducting MLR with many factors because, by design, factors were uncorrelated. However, the inclusion of both raw habitat variables and factors had the potential to cause variance inflation, creating significant uncertainty in MLR coefficient estimates and significance. Because factors were calculated using the complete sediment chemical concentration dataset, collinearity between factors could arise when developing models for benthic response variables measured at a subset of the full dataset (e.g., toxicity test endpoints and HBI). This correlatedness was likely artificial.

The variance inflation factor (VIF), an indicator of the collinearity of two variables in a MLR, was calculated for the full and limited models, which included PCs and raw habitat variables.[35] A VIF of five or greater was generally considered to be significant, and to suggest that correlation between explanatory variables had affected the model. To address collinearity, path analysis was used to identify the less important

[35] The R function "vif" from the "car" package (Fox and Weisberg 2011) was used to calculate the VIF.

variable(s) among those contributing to collinearity, so that these variable(s) could be removed from the final MLR model.

Path analysis was a means of investigating potential "causal" relationships between competing explanatory factors in a MLR.[36] Theoretical "paths" between variables in a MLR were typically constructed by the statistician using a conceptual model and institutional knowledge. Explanatory variables could be interrelated as well as related to the response variable. Ultimately, the output of path analysis was a set of coefficient estimates, among other things (e.g., Z-test p-values), that showed the importance of each explanatory variable for predicting the response variable. Based on the path analyses herein, collinear variables with relatively low (and/or insignificant) path coefficients were removed from the final MLR models. When multiple chemical factors had elevated VIF values (i.e., for sediment toxicity of HBI models with reduced datasets), higher order factors were preferentially removed, because they described smaller fractions of the variance associated with chemical concentration data. Correlation matrices were used to inform the construction of paths between variables.[37]

MLR Model Comparison

MLR models for each benthic response variables were compared using various statistics. These statistics evaluated the various strengths and weaknesses of the models, and aided in ranking the models in terms of goodness-of-fit and uncertainty (e.g., resulting from the classic bias-variance tradeoff related to overfitting or underfitting).[38] High model variance was a concern for the full model, which included more chemical factor parameters than the other models, some of which explained little of the overall variance associated with sediment chemical concentration data. High model bias was a concern for the habitat-only models, which did not include potentially important chemical factors. Unless otherwise noted, all statistics were calculated in R using built-in functions and packages.

[36] The R function "sem" from the "lavaan" package (Rosseel 2012) was used to conduct path analyses. The analyses were based, in part, on a correlation matrix generated using the "rcorr" function from the "Hmisc" package (Harrell 2016). Path analysis attempts to evaluate causality by controlling for partial correlations, but it is not truly a test of causality; results are ultimately based on correlation.

[37] Intercorrelated factors, when observed, were said to be correlated in the model rather than regressed on one another because the factors were, by design, independent. In the case of correlated factors and habitat variables, the factors were said to be regressed on the habitat variables, assuming that chemical concentrations were dependent upon the physical characteristics of sediment.

[38] Overfitting in MLR results from the inclusion of many variables (some of which are superfluous), the outcome being a model that is sensitive to random fluctuations in a response variable. Although this can result in high-performing models based on statistics such as r^2, overfit models tend to perform very poorly when presented with new data or during cross-validation (i.e., high variance in the bias-variance tradeoff). Underfitting occurs when important explanatory variables are not included, resulting in models with poor fit (i.e., high bias).

The following statistics, described in more detail below, were used to rank MLR models:

- Akaike's information criterion (AIC)
- Bayesian information criterion (BIC)
- Adjusted r^2
- Predicted residual error sum of squares (PRESS)
- Predicted r^2
- F-test result (based on sequential "type I" sum of squares)
- Number of samples with Cook's distance (D) > $\frac{4}{n}$, where n = sample size
- Number of samples with DFFITS > $2 * \sqrt{\frac{p}{n}}$, where p = number of explanatory variables and n = sample size

The AIC and BIC statistics were relative measures of model fit. Their calculation accounted for overfitting by penalizing each statistic for the number of variables included in a model. The BIC tended toward the selection of smaller models, whereas the AIC tended toward the selection of larger models (Dziak et al. 2012). Taken together, the AIC and BIC could bracket an appropriately sized MLR model. As a general rule, a decrease of 10 or more in AIC or BIC between models indicated a notable improvement in model fit.

Adjusted and predicted r^2 values were used to evaluate model fits.[39] Adjusted r^2 was similar to the standard r^2 value commonly evaluated for linear models, but it was penalized for the number of coefficients to account for overfitting. Penalization of r^2 was useful because r^2 always increased as more explanatory variables were added to an MLR, regardless of the importance of the added factors. The predicted r^2 value differed from the adjusted r^2 value in that it accounted for overfitting using a cross-validation approach rather than penalization. PRESS, which was used to calculate the predicted r^2, was calculated by 1) removing a single sample from the MLR dataset, 2) fitting the MLR, 3) estimating the squared residual error of the model, 4) returning the removed sample and removing a different sample, 5) repeating steps 2 through 4 until each sample had been removed, and 6) summing the squared residual errors across all iterations of step 3. By accounting for minor changes in a dataset (using cross-validation), PRESS and the predicted r^2 provided an indication of the

[39] Although commonly reported as a traditional measure of model fit, coefficient of multiple determination (r^2) values are generally not indicative of the strength of actual relationships which may not follow the assumptions of fitted models. Low r^2 values (such as those derived through correlation analyses in the BERA) are not necessarily an indicator of a weak relationship, and high r^2 values are not necessarily indicative of a correct model.

performance of a model in general, rather than specific to a single dataset. The functions used to calculate PRESS and predicted r^2 were developed in R.

Analysis of variance (ANOVA) could be used to compare the fit of "nested" models, or those that contained only a subset of the largest model being compared. For example, the null, habitat-only, limited chemistry, and limited combined MLR models were all nested with the full model. Because "type I" sum of squares was used to conduct ANOVA, the results of the F-test were sequential, such that a significant result ($p < 0.05$) indicated whether the habitat-only model performed better than the null model, whether the limited combined model performed better than the habitat-only model, or whether the full combined model performed better than the limited model.[40] A sequential comparison could not be made between models with only habitat variables or only chemistry variables, because they were not nested models of increasing complexity. Variables were not simply added; rather, each model was based on mutually exclusive explanatory variable subsets. In order to determine if habitat variables contributed significantly to model performance after first accounting for chemical factors, a second ANOVA was conducted by testing sequentially the null, limited chemistry, and limited combined models; the limited combined model was the same as the limited chemistry model but with habitat variables added.

Cook's D and DFFITS were measures of the influence of single samples over a regression (i.e., model-predicted values) and were indicative of potential outliers. The Cook's D threshold of $4/n$, described above, was suggested by Bollen and Jackman (1990), and the DFFITS threshold of 2 times $\sqrt{p/n}$ was suggested by Belsley et al. (1980).

Although commonly reported as a traditional measure of model fit, coefficients of multiple determination (r^2) values are generally not indicative of the strength of actual relationships, which may not follow the assumptions of fitted models. Low r^2 values (such as those derived through correlation analyses in the BERA) are not necessarily an indicator of a weak relationship, and high r^2 values are not necessarily indicative of a correct model.

4.1.3 T20 and T50 sediment chemistry screen

LPRSA sediment chemistry data were compared to logistic regression model-based screening criteria for select chemicals (i.e., those for which such criteria exist) for the purpose of assigning weights to the sediment chemistry LOE in the WOE analysis (Section 6.1 of the BERA main text). The T20 and T50 values described by Field et al. (2002) were used to evaluate LPRSA chemistry data (Appendix B, Table B7-1). LPRSA locations with at least one chemical concentration exceeding a T20 value were assigned a weight of 0.5 in the WOE analysis, and locations with at least one

[40] A model "perfomed" better in the case of ANOVA if the larger model significantly reduced the residual sum of squares relative to the smaller model.

concentration exceeding a T50 value were assigned a weight of 1.0 in the WOE analysis.

4.1.4 Uncertainty analyses

Uncertainty analysis methods are described in subsections below. The results of uncertainty analyses are described in Section 4.3 and subsections.

4.1.4.1 Predictive accuracy of T20 and T50 criteria

LPRSA toxicity test results were used to evaluate the reliability of T50 values (Field et al. 2002) to predict toxicity of LPRSA sediment relative to reference conditions. This evaluation was accomplished by calculating several reliability statistics that are common to contingency tables. Reliability statistics are used to quantify the ability of a classification scheme to predict categorical results (e.g., toxic or not toxic) (James et al. 2013). Sediment chemistry data are often compared to sediment quality guidelines (e.g., T50 values) in an effort to classify sediment samples as being either likely toxic or likely non-toxic to aquatic species. This comparison can be evaluated using reliability statistics when empirical sediment toxicity test data are co-located with sediment chemistry data, as in the LPRSA. The actual toxicity of LPRSA sediments (relative to reference conditions) is described in Appendix B (Tables B3 and B4) and summarized in Table 3-11. T50 exceedances are described in Section 4.3.1 and Appendix B (Table B7-1).

The following reliability statistics were calculated to assess the accuracy of T50 values:

- Number of true positives, false positives, true negatives, and false negatives
- Type I and II error rates (and total error rate)
- Sensitivity and specificity
- Precision and false discovery rate

A true positive result is defined as LPRSA sediment toxicity that was outside the reference envelope and that was predicted to be toxic due to an exceedance of any T50 at the same location. A false positive is defined as an LPRSA location with actual sediment toxicity within the reference envelope (i.e., not toxic) but that was predicted to be toxic due to an exceedance of a T50 at the same location. True and false negatives represent analogous scenarios wherein a sediment sample was predicted to be non-toxic. The type II error rate is the number of false positives divided by the total number of samples, and the type I error rate is the number of false negatives divided by the total number of samples. The total error rate is the sum of the type I and type II error rates and therefore reflects the percentage of results misclassified by the T50.

Sensitivity and specificity can be defined as the probability that the T50 will predict toxicity when there is actually toxicity (sensitivity) or predict no toxicity when there is actually no toxicity (specificity). Sensitivity and specificity are each calculated on a scale between 0 and 1, with 1 being perfect sensitivity or specificity. A perfectly

accurate classifier (e.g., T50) would have both specificity and sensitivity values equal to 1. Sensitivity is calculated as the number of true positives divided by the sum of actually positive results (i.e., true positives and false negatives). Specificity is calculated similarly, as the number of true negatives divided by the number of actually negative results (i.e., true negatives and false positives).

Precision and the false discovery rate are measures of the probability of obtaining a false positive result. A precision value equal to 1 indicates that a classifier (e.g., T50) always correctly predicts toxicity, whereas a false discovery rate equal to 1 indicates that the predictor always falsely predicts toxicity. Precision is calculated as the number of true positives divided by the sum of predicted positives (i.e., true positives and false positives). The false discovery rate is equal to 1 minus precision.

4.1.4.2 SEM-AVS analysis

Extensive testing of the fate, transport, and toxicity of metals in sediment over a period of decades has resulted in the development of the EqP as a paradigm for metals toxicity in sediment. The EqP paradigm is based on findings that sediment toxicity tends to be better predicted from pore water concentrations than from total, dry weight concentrations of metals in bulk sediment (USEPA 2005b). The toxicities of metals are driven by their respective bioavailabilities, and dissolved metals are more bioavailable than metals bound to particulate material or precipitates (Di Toro et al. 2001; DiToro et al. 2005; EPA 2003; Paquin et al. 2002; Santore et al. 2001; Wood 2012).

Metal sulfides are a primary example of insoluble metal complexes that can be formed in sediment and sediment pore water (Newman 1998). The cold acid-extraction method is used to determine the level of sulfides in sediment; sulfides extracted using this method are called AVS, and the divalent metals that dissolve during the analytical procedure are called SEM (Allen et al. 1991). The difference between the molar concentrations of ΣSEM and AVS has been used to predict whether several divalent metals (i.e., cadmium, copper, lead, nickel, silver, and zinc)[41] in sediment are non-toxic (USEPA 2005b). This follows from the EqP paradigm, in that the concentration of soluble metals does not exceed the complexation capacity of sulfides when the molar concentration of ΣSEM is less than the molar concentration of AVS (i.e., ΣSEM – AVS < 0 µmol/g), resulting in non-bioavailable metals.

A key uncertainty associated with the SEM-AVS method is that analytical measurements of SEM-AVS tend to be irreproducible among laboratories, owing to the volatile nature of the analytes (Hammerschmidt and Burton 2010). The SEM-AVS analyses for LPRSA SQT sediments were conducted by a single laboratory (Columbia Analytical Services in Kelso, Washington), thereby reducing this uncertainty.

[41] Additional SEM exist; however, their relationships to laboratory toxicity test data are not well characterized in the literature.

When the molar concentration of ΣSEM in bulk sediment exceeds the molar concentration of AVS (i.e., ΣSEM-AVS > 0 µmol/g), the sediment *might* be toxic due to the bioavailability of metals. In order to estimate potential metal toxicity in such cases, it is first necessary to consider the OC content of sediment, because OC also sorbs metals and limits the bioavailability of metals not bound to AVS (i.e., the excess SEM) (USEPA 2005b). The influence of OC is accounted for by normalizing the excess SEM concentration to the fraction of OC in the sediment (i.e., [ΣSEM-AVS]/f_{OC}). Based on an evaluation of sediment toxicity data versus OC-normalized excess SEM concentrations, the following concentrations were derived for predictions of no toxicity expected, uncertain toxicity, and a high likelihood of toxicity (USEPA 2005b):

- No toxicity expected – (ΣSEM-AVS)/f_{OC} ≤ 130 µmol/g OC
- Uncertain toxicity[42] – (ΣSEM-AVS)/f_{OC} > 130 and ≤ 3,000 µmol/g OC
- High likelihood of toxicity – (ΣSEM-AVS)/f_{OC} > 3,000 µmol/g OC

Although there were no SEMs identified as COPECs in the SLERA,[43] several SEMs were assessed as part of the sediment chemistry LOE and compared with measured sediment toxicity and benthic community metrics. The results from this analysis are presented in Section 4.3.2.

4.1.4.3 Equilibrium partitioning-based analysis of 34 PAHs

In response to discussions with USEPA Region 2, CPG has provided an analysis of the potential for PAH toxicity using the sum of 34 PAHs. The 34-PAH-sum approach to predicting PAH toxicity is described by USEPA (2003) guidance. The 34-PAH-sum approach differs from the other analyses of sediment chemistry in this ERA, which are based on a summed, dry weight concentration of 16 PAHs.[44]

Similar to the EqP analysis of metals (Section 4.1.5.2), USEPA provides guidance on the analysis of 34 PAH mixtures using an EqP approach (USEPA 2003). The EqP paradigm holds that PAHs partition among OC, interstitial water, and benthic invertebrate tissues (among other substrates). Following USEPA (2003) guidance, partitioning coefficient values (i.e., Kow and Koc) and toxic equivalency factors are used to calculate EqP-based PAH sums (in toxic units) (referred to hereafter as ΣESB TUs) for each sediment sample. If a ΣESB TU exceeds 1, then the potential exists for toxic impacts on sensitive benthic invertebrate species.

In discussion with USEPA Region 2, it was agreed that CPG would provide an analysis of uncertainty associated with the use of sums of 16 PAHs rather than toxic unit-based 34-PAH sums. However, a direct comparison using USEPA (2003) methods

[42] Within this range of concentrations, toxicity varies substantially (USEPA 2005b).

[43] No New Jersey Department of Environmental Protection (NJDEP) ecological screening criteria for SEMs are available, so SEMs could not be identified as COPECs using these criteria in the SLERA.

[44] The method for summing PAHs in the BERA is described in the *Data usability and data evaluation plan for the Lower Passaic River Study Area risk assessments* (Windward and AECOM 2015).

is neither possible nor relevant given the discussion of sediment chemistry provided in this LOE. Instead, the toxic unit-based 34-PAH sums are evaluated on their own, similarly to SEM-AVS, in Section 4.3.3.

4.1.4.4 Quantitative analysis of the uncertainty associated with sediment chemistry LOE criteria

An analysis was conducted to address uncertainties associated with using the T20 and T50 values to define the sediment chemistry LOE. Uncertainties in those values are quantified in Section 4.3.1. The analysis described in the following subsections represents an alternative approach to classifying LPRSA SQT samples as likely toxic or likely nontoxic based on measured sediment chemical concentrations.

Mean-quotient Threshold Comparison

LPRSA sediment chemistry data were evaluated using a mean-quotient approach and reference area sediment toxicity test data. Sediment chemical concentration data from Jamaica Bay and Mullica River/Great Bay were used to calculate mean ERM quotient (mERMq) values (for each reference area), and sediment concentration data from the area above Dundee Dam were used to calculate mPECq values. These reference area mean-quotients were then compared to co-located sediment toxicity data (as a percent of negative control results), and low and high mERMq and mPECq thresholds were selected.

The low mERMq threshold for each reference area dataset was set as the highest mERMq value below which *A. abdita* survival was entirely ≥ 80% of the negative control result (i.e., an "acceptable" estuarine reference sample), and the low mPECq threshold was set as the highest mPECq value below which *C. dilutus* or *H. azteca* survival was entirely ≥ 75% of the negative control result (i.e., an "acceptable" freshwater reference sample). The lower (i.e., more protective) mPECq value between the two freshwater toxicity test species was selected as the low mPECq threshold.[45] The high mERMq threshold was set as the lowest mERMq value above which *A. abdita* survival was entirely < 80% of negative control, and the high mPECq threshold was set as the lowest mPECq value above which *C. dilutus* or *H. azteca* survival was entirely < 75% of negative control. Again, the lower mPECq value between the two freshwater toxicity test species was selected as the high mPECq threshold. These values are presented in Appendix B, Table B7-2.

Sediment chemical concentration data from the LPRSA (i.e., mERMq and mPECq values) were screened against the mERMq and mPECq reference thresholds to categorize potential chemical risks. LPRSA values below the low thresholds were categorized as having negligible potential to cause toxicity (0.0 weight in the

[45] The values of 80 and 75% of control for estuarine and freshwater sediment toxicity test results are the same values used to screen the reference datasets used to characterize benthic risk (as described in Section 2.3.1 and Appendix B, Tables B3-3 and B4-3).

quantitative analysis of uncertainty). LPRSA values exceeding the high thresholds were categorized as having high potential to cause toxicity (1.0 weight in the quantitative analysis of uncertainty). LPRSA values exceeding the low threshold but not exceeding the high threshold were categorized as having an unclear potential to cause toxicity (0.5 weight in the quantitative analysis of uncertainty). These values are presented in Appendix B, Table B7-3.

This approach to evaluating LPRSA sediment chemistry is based on site-specific effects data rather than literature-based analyses (e.g., T20 and T50 values). Sections 4.1.5.1 and 4.1.5.2 describe uncertainties associated with the two chemical LOE approaches.

Predictive Accuracy of Mean-quotient Threshold Comparison

LPRSA toxicity test results were used to evaluate the reliability of mean-quotient values (Appendix B, Table B7) to predict the toxicity of LPRSA sediment relative to reference conditions. This evaluation was conducted as described in Section 4.1.5.1 for T20 and T50 values, but using the low and high mean-quotient threshold values derived for freshwater and estuarine reference areas in place of the T20 and T50 values. The comparison of LPRSA chemistry data to mean-quotient thresholds is presented in Appendix B, Table B7-3.

4.2 RESULTS

4.2.1 Bivariate correlation analysis

Results for the bivariate Spearman rank correlation analyses conducted for the LPRSA SQT dataset are summarized in Tables 4-2 and 4-3, and the complete results are provided in Appendix B (Table B1-6).

Table 4-2. Spearman rank correlation coefficients for significant correlations, site-wide LPRSA (RM 0 to RM 17.4)

Chemical	*C. dilutus* Weight	*C. dilutus* Biomass	*C. dilutus* Survival	*H. azteca* Weight	*H. azteca* Biomass	*H. azteca* Survival	*A. abdita* Survival	Abundance (per m^2)	Taxa Richness	Shannon-Wiener H'	Pielou's J'	SDI	HBI[a]
Lead	-0.27	--	--	-0.35[b]	-0.43[b]	-0.41[b]	--	--	-0.36[a,b]	-0.36[a,b]	--	-0.30	0.65
Mercury	-0.30	--	--	-0.27	-0.31	-0.31	--	-0.26	-0.33	-0.26	--	-0.22	0.61
Zinc	-0.28	--	--	-0.23	-0.31	-0.33	-0.42	--	-0.36[b]	-0.34	--	-0.26	0.63
Total HPAHs	--	--	--	--	-0.23	-0.24	--	--	--	--	--	--	--
Total LPAHs	--	--	--	--	--	--	--	--	--	--	--	--	--
Bis-(2-ethylhexyl) phthalate	--	--	0.30[c]	-0.35	-0.35[b]	-0.28	--	-0.30	-0.51[b]	-0.36[b]	--	-0.26	0.49
Total PCB Congeners	--	--	--	-0.28	-0.27	-0.28	--	-0.30	-0.47	-0.42	--	-0.38	0.48
Total DDx	--	--	--	--	--	--	--	-0.34	-0.55[b]	-0.47[b]	--	-0.39[b]	0.51
TOC[d]	--	--	--	-0.22	-0.26	-0.25	--	-0.24	-0.51	-0.52	-0.22	-0.44	--
Percent fines[d]	--	--	--	--	--	--	0.43	-0.29	-0.32	-0.20	--	--	--

[a] Positive correlations with HBI are suggestive of a chemical response.

[b] Correlation is significant after Bonferroni correction (corrected alpha = 0.0005).

[c] Correlation is positive and therefore not suggestive of chemical response.

[d] TOC and percent fines are included for context but are included in summaries made in the text (i.e., percent of significant correlations).

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
HBI – Hilsenhoff Biotic Index
HPAH – high-molecular weight polycyclic aromatic hydrocarbon
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
LPRSA – Lower Passaic River Study Area
PCB – polychlorinated biphenyl
RM – river mile
SDI – Swartz's dominance index
TOC – total organic carbon
total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT and 4,4′-DDT)

Table 4-3. Spearman rank correlation coefficients for significant correlations, tidal freshwater zone (RM 13 to RM 17.4)

Chemical	*C. dilutus* Weight	*C. dilutus* Biomass	*C. dilutus* Survival	*H. azteca* Weight	*H. azteca* Biomass	*H. azteca* Survival	Abundance (per m^2)	Taxa Richness	Shannon-Wiener H'	Pielou's J'	SDI	HBI[a]
Total PAHs	--	--	--	--	--	--	--	--	--	--	--	--
Bis-(2-ethylhexyl) phthalate	--	--	--	-0.62	-0.52	--	--	--	--	--	--	0.49
Phenol	--	--	--	--	--	--	0.51[b]	--	--	--	--	--
Total PCB Congeners	--	--	--	--	--	--	--	--	--	--	-0.51	0.48
Total chlordanes	--	--	--	--	--	--	--	--	--	--	--	--
Total DDx	--	--	--	--	--	--	--	-0.48	-0.48	--	--	0.51
TOC[c]	--	--	--	--	--	--	--	--	-0.48	--	--	--
Percent fines[c]	--	--	--	--	--	-0.52	0.51[b]	--	--	--	--	--

Note: No correlation reported in this table was significant after Bonferroni correction (corrected alpha = 0.0007).

[a] Positive correlations with HBI are suggestive of a chemical response.

[b] Abundance is positively correlated, which may suggest a chemical response.

[c] TOC and percent fines are included for context but are included in summaries made in the text (i.e., percent of significant correlations).

DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
HBI – Hilsenhoff Biotic Index
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
RM – river mile
SDI – Swartz's dominance index
TOC – total organic carbon
total DDx – sum of all six DDT isomers (2,4′-DDD, 4,4′-DDD, 2,4′-DDE, 4,4′-DDE, 2,4′-DDT and 4,4′-DDT)

As can be seen in Table 4-2, there are several significant and relevant correlations between benthic response variables and paired sediment chemistry data (as well as TOC and percent fines). Of the 104 correlations tested,[46] 50 were significant (48%). Pielou's J' and *C. dilutus* biomass and survival were not correlated with any chemicals, and total low-molecular-weight polycyclic aromatic hydrocarbons (LPAHs) were not correlated with any benthic responses.[47] The strongest correlations were observed between metals and HBI. Weaker correlations were observed for *H. azteca* toxicity test endpoints, *C. dilutus* growth, abundance (per m^2), and SDI. Apart from correlations between metals and HBI, coefficients of variation (squared r-values, which correspond with the percentage of variance explained by the bivariate relationship) are quite low ($r^2 \leq 0.31$) for all significant relationships. After Bonferroni correction, the number of significant correlations decreased, as expected.

Only 9 of 72 tested correlations[48] (10%) were significant for the tidal freshwater dataset (Table 4-3). Although many of the correlations for that dataset are weak (i.e., low r-value) (Appendix B, Table B1), the r-values are not uniformly low relative to r-values from analogous correlations tested with site-wide data. This suggests that the correlation analyses of tidal freshwater data were limited by the small freshwater dataset sample size (relative to the site-wide dataset). The tidal freshwater sample size was 18 for benthic community metrics and 19 for toxicity test data, compared to (for example) 97 and 98 for benthic community metrics and *H. azteca* toxicity test endpoints, respectively, in the site-wide dataset.[49] Regardless, coefficients of variation among significant correlations of paired freshwater data were generally weak ($r^2 < 0.3$), excepting only the correlation between *H. azteca* growth and bis(2-ethylhexyl) phthalate ($r^2 = 0.39$).

Correlations between chemicals, presented in Appendix B, Table B1, indicate fairly strong relationships between metals and non-PAH organic chemicals. Total LPAHs are particularly uncorrelated with other chemicals in the LPRSA.

Given that there are many significant correlations between chemistry and benthic responses site wide, there is reason to believe that chemistry and benthic responses are related in some way, although bivariate correlations are weak. This indicates that single chemical-benthic response relationships are unreliable for predicting ecological risks to benthic invertebrates throughout the LPRSA based on sediment chemistry data alone.

[46] The count of 104 excludes correlations conducted with TOC and percent fines.
[47] Pielou's J' was only significantly correlated with TOC.
[48] The count of 72 excludes correlations conducted with TOC and percent fines.
[49] Other sample sizes are as follows: *C. dilutus* toxicity test endpoints (n = 71), *A. abdita* survival (n = 27), and HBI (n = 18). HBI was only calculated for freshwater locations, so correlation results are identical between the two datasets.

4.2.2 Multivariate analysis

The results of the multivariate analysis are detailed in Appendix B, Table B2, and discussed in this section. The results based on analyses of the two multivariate datasets (Method 1 and Method 2) are described separately first (Sections 4.2.2.1 and 4.2.2.2) and then compared (Section 4.2.2.3).

4.2.2.1 Method 1 results

Principal Component Analysis

Eigenvalues from the Method 1 PCA indicated that the vast majority (86%) of the total variance in sediment chemical concentration data could be explained by the first 2 (of 30) PCs, but that 7 PCs were required to explain 95% of the total variance. Only 3 PCs were required to explain 90% of the total variance.

Exploratory Factor Analysis

Loading values indicated that the first three factors (referred to as F1, F2, and F3) were generally associated with:

- F1 – all PAHs; some metals (e.g., copper, mercury, lead, and zinc) to a moderate extent; and alpha-chlordane, Dieldrin, and total DDx to a moderate extent.
- F2 – primarily all metals (particularly arsenic, chromium, nickel, and silver); total DDx and hexachlorobenzene to a lesser extent
- F3 – primarily organochlorine pesticides

Factors 4, 5, 6, and 7 were weakly associated with a number of chemical variables. The factor scores from the Method 1 EFA are provided in Appendix B, Table B2-3.

Multiple Linear Regression

The results of MLR analyses are presented in Table 4-4, which includes an indication of the model type, relevant statistics (Section 4.1.2.3, MLR Model Comparison), the "best" model selected, modeling development notes, and rationale for model selection. Caveats are related to specific treatments of model datasets, such as dropping variables to reduce VIF and/or low or non-significant path coefficients (Section 4.1.2.3, Collinearity and Path Analysis), or transforming response variables to address non-normality, trends, or heteroscedasticity of residuals (Section 4.1.2.3, Checking MLR Assumptions). Included in the model notes is an indication of which variables were significant[50] (as well as the implication of significance) when included in the full model, even if the full model was not selected as "best." This investigation is meant to evaluate which coefficients are important in terms of sign (i.e., positive or negative relationship to the response variable) and (absolute) magnitude (i.e., degree of

[50] Significance is based on a t-test, alpha = 0.05. A significant t-test result indicates that the slope parameter for the given variable is different from 0, suggesting a relationship.

influence over the model, when all other explanatory variables are held constant). Chemical factor coefficients and habitat variable coefficients are directly comparable because they are in the same units (standard deviation).

Table 4-4. MLR results, Method 1

Endpoint	Model	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant ($p < 0.5$)?[a]		No. of Samples with Cook's D > 4/n	No. of Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
Abundance (per m^2)	null	2	468	474	--	124	-0.0103	--		12	105	The full model is significantly better than other models according to the F-test results, but the relevance of this result is dubious given the quality of the model fit. The inclusion of habitat improves the null model but not the limited chemistry model, suggesting that chemistry is marginally more important than habitat for predicting abundance. Total fines, F3, and F7 are significant variables in the full model (t-test, $p < 0.05$). F7 is essentially uninterpretable. Total fines has the largest coefficient among significant variables. F3 and F7 have similar coefficient values.
	habitat-only	4	459	472	0.0520	120	0.0193	yes		12	11	
	limited combined	8	453	480	0.0980	116	0.0500	yes	no	11	4	
	limited chemistry	6	455	475	0.0800	116	0.055	yes		11	5	
	full	11	448	484	0.136	113	0.0749	yes		14	6	
Richness	null	2	1,450	1,456	--	18,500	-0.0103	--		10	80	Based on Q-Q plots, the normality of richness model residuals is suspect; log-transformation of richness does not address this uncertainty, so the raw data are modeled. Based on model statistics, the limited chemistry model is best, although it appears to be influenced by the more extreme values (based on Cook's D and DFFITS). Habitat variables do not improve the model, nor does the full set of chemical factors. The limited chemistry model performs better than the habitat-only model based on AIC, BIC, adjusted and predicted r^2, and PRESS. F1, F2, and F3 are all significant factors when included in the full model (t-test, $p < 0.05$); F1 and F2 have similar coefficients, and F3 is approximately 2 times larger then F1 and F2.
	habitat-only	4	1,435	1,449	0.0790	17,357	0.0521	yes		12	15	
	limited combined	8	1,379	1,406	0.322	13,161	0.281	yes	no	19	15	
	limited chemistry	6	1,376	1,395	0.327	12,850	0.298	yes		22	15	
	full	11	1,383	1,419	0.320	13,633	0.255	no		20	12	
Shannon-Wiener H'	null	2	328	335	--	60.6	-0.0103	--		13	99	Model statistics (i.e., AIC, BIC, PRESS, and predicted r^2) indicate that the limited chemistry model is best. The addition of habitat variables does not improve the chemistry model, nor does adding the full set of chemical factors. F2 and F3 are significant factors when included in the full model (t-test, $p < 0.05$); the coefficients for these factors are similar.
	habitat-only	4	302	315	0.135	54.0	0.0994	yes		9	11	
	limited combined	8	256	282	0.328	43.8	0.269	yes	no	18	9	
	limited chemistry	6	256	275	0.322	42.8	0.287	yes		16	9	
	full	11	260	296	0.325	45.6	0.239	no		18	9	
Pielou's J'	null	2	-169	-163	--	4.7	-0.0103	--		10	97	The full model is best based on the higher r^2 values and slightly lower values for other statistics, but this result is very uncertain due to the poor fit of the model. Adding habitat significantly improves the limited chemistry model. TOC, total fines, F2, F3, and F7 are significant factors in the full model (t-test, $p < 0.05$). F7 is essentially uninterpretable. Total fines is positively related to J′. The coefficients for total fines, F2, and F7 are largest (and similar to one another).
	habitat-only	4	-172	-159	0.0230	4.7	-0.00438	yes		14	7	
	limited combined	8	-173	-147	0.0490	4.7	-0.00263	yes	yes	15	10	
	limited chemistry	6	-173	-153	0.0370	4.7	0.00433	yes		15	12	
	full	11	-182	-146	0.105	4.5	0.0432	yes		13	11	
SDI (log)	null	2	-36	-30	--	9.5	-0.0103	--		18	94	Q-Q plots suggest that model residuals are not normal, and residual-fitted value plots suggest heterskedasticity of residuals. Log-transformation of SDI values sufficiently addresses both uncertainties. The limited chemistry model is best, as suggested by the AIC, BIC, PRESS, and predicted r^2 statistics. The addition of habitat variables does not improve the limited chemistry model, nor does adding the full set of chemical factors. Total fines, F2, and F3 are significant factors in the full model (t-test, $p < 0.05$). Total fines is positively related to SDI. Total fines is positively related to SDI. F2 has the largest coefficient, but the coefficients for F3 and total fines are only slightly lower than that for F2 (within a factor of 2).
	habitat-only	4	-56	-43	0.104	8.7	0.0703	yes		9	12	
	limited combined	8	-87	-60	0.250	7.6	0.191	yes	no	17	10	
	limited chemistry	6	-87	-67	0.242	7.4	0.205	yes		16	8	
	full	11	-85	-49	0.256	7.7	0.172	no		16	10	

Endpoint	Model	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant ($p < 0.5$)?[a]		No. of Samples with Cook's D > 4/n	No. of Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
HBI	null	2	170	175	--	41.5	-0.0272	--		4	43	Models are based on a small subset of data (n = 75, reduced from n = 205). VIF values are elevated for total fines and F2; path analysis indicates that F2 is more important than total fines for explaining HBI. Total fines are removed from models to reduce variance inflation (model uncertainty). Removing total fines does not substantially improve model fit. Diagnostic plots indicate that residuals are not normal for the habitat-only model, and that some extreme values may be influencing regressions. Log-transformation of HBI is not useful due to the small range of possible HBI values (0–10). The same problem is not as pronounced for the other models. The full model is the best model based on its higher r^2 values and similar AIC and BIC values, even though the model is more complex. The predicted r^2 value is highest for the full model, suggesting that it is not overfit. Also, the F-test results indicate that the addition of both habitat variables and the full chemical factor list significantly improves the model. F1, F2, F3, and F7 are significant factors in the full model (t-test, $p < 0.05$). F7 is essentially uninterpretable. The F1 and F3 coefficients are greatest (and similar), and the F2 and F7 coefficients are approximately half of those for F1 and F3.
	habitat-only	3	166	173	0.0730	39.1	0.0315	yes		5	0	
	limited combined	7	99	115	0.638	18.2	0.550	yes	yes	8	3	
	limited chemistry	6	97	111	0.641	17.8	0.559	yes		7	3	
	full	10	93	116	0.678	16.7	0.587	yes		9	4	
C. dilutus survival	null	2	-85	-80	--	2.2	-0.0216	--		4	59	Models are based on a small data subset (n = 94, reduced from n = 205). VIF values are elevated for F1 and F2 (and to a lesser extent, total fines and F3); path analysis indicates that total fines and F2 (as well as F4 and F7) are not significant variables, and that F3 is positively related to survival. Total fines, F3, F4, and F7 are removed from models to address inflation across several variables. F1 and F2 are retained to account for the majority of variance associated with chemical variables. Diagnostic plots indicate that there are several extreme values that are likely influencing regressions. The limited chemistry model is best, in that it has AIC, BIC, PRESS, predicted r^2, and extreme values similar to those of other models, but with a lower number of variables. Habitat variables (in this case, only TOC) improve neither the null model nor the limited chemistry model. Due to the very poor fit of all models, the best result is highly uncertain. F1 and F6 are significant factors in the full model (t-test, $p < 0.05$). F6 is essentially uninterpretable.
	habitat-only	3	-86	-78	0.0190	2.3	-0.0898	no		4	1	
	limited combined	5	-94	-81	0.120	2.3	-0.0684	yes	no	4	1	
	limited chemistry	4	-96	-86	0.130	2.2	-0.0247	yes		6	2	
	full	7	-97	-80	0.169	2.2	-0.0274	yes		5	2	
C. dilutus biomass	null	2	-25	-20	--	4.1	-0.0216	--		4	52	All models are based on a small data subset (n = 94, down from n = 205). VIF values are elevated for F1 and F2 (and, to a lesser extent, total fines, F3, F4, and F5); path analysis indicates that total fines, F3, and F4 are not significant, and that F3 has a positive relationship with biomass. F3 and total fines are removed from models, which has an adverse impact on model fits but reduces model uncertainty. Based on model statistics, the limited combined model is best. The F-test result for that model (comparison to limited chemistry) indicates that the addition of TOC significantly improves the model, although model statistics (excepting r^2 values) suggest that the models are quite similar. Adding the full set of chemical factors does not improve on the limited combined model. Due to the poor fit of all models, the best result is highly uncertain. TOC, F1, and F2 are significant factors in the full model, although TOC is positively related to biomass. Coefficients for each variable are similar to each other (within a factor of 2).
	habitat-only	3	-25	-17	0.00800	4.2	-0.0448	no		6	6	
	limited combined	6	-42	-27	0.202	3.6	0.119	yes	yes	6	5	
	limited chemistry	5	-37	-24	0.148	3.7	0.0890	yes		6	5	
	full	9	-40	-17	0.207	3.6	0.0981	no		6	4	
H. azteca survival	null	2	-12	-7	--	6.3	-0.0167	--		9	68	Models are all based on a small data subset (n = 121, down from n = 205). Q-Q plots indicate that the normality of residuals is suspect. Simple transformations (i.e., logarithmic or square-root) do not sufficiently address the issues associated with residuals. VIF values are elevated for F1 and F2 (and, to a lesser extent, total fines); path analysis indicates that total fines, F3, and F4 are not significant, and that F3 has a positive relationship with biomass. Correlation tables suggested that F4 is correlated with several of the inflated variables as well. Removing F4 from models decreases VIF to an acceptable level (< 5) for all variables while retaining the more important factors of F1, F2, and F3. Based on model statistics, the limited combined model is best. The F-test result for that model (comparison to limited chemistry) indicates that the addition of habitat variables significantly improves the limited chemistry model; other statistics also suggest that the limited combined model is best. Adding the full set of chemical factors does not improve the limited combined model. TOC, total fines, F2, and F3 are significant factors in the full model, although both total fines and F3 have positive relationships with *H. azteca* survival. F2 and total fines have the largest coefficients; the coefficients (absolute) for F3 and TOC are approximately 2.5 times lower than that for F2.
	habitat-only	4	-30	-19	0.151	5.5	0.115	yes		9	4	
	limited combined	7	-62	-42	0.362	4.2	0.319	yes	yes	8	3	
	limited chemistry	5	-40	-26	0.221	5.1	0.180	yes		8	4	
	full	10	-57	-29	0.349	4.5	0.280	no		8	3	

Endpoint	Model	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant ($p < 0.5$)?[a]		No. of Samples with Cook's D > 4/n	No. of Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
H. azteca biomass	null	2	-23	-18	--	5.8	-0.0167	--		7	59	All models are based on a small data subset (n = 121, down from n = 205). VIF values are elevated for F1 and F2 (and, to a lesser extent, total fines). Correlation analysis suggests that other factors (e.g., F5) are also correlated with F1, F2, and total fines. Path analysis indicates that F5, F6, and F7 are not significant. Removing F5 from models decreases VIF to an acceptable level (< 5) for all variables while retaining the more important factors, F1, F2, and F3. Based on model statistics, the limited combined model is best. The F-test result for that model (comparison to limited chemistry) indicates that the addition of habitat variables significantly improves the limited chemistry model; other statistics (e.g., PRESS and predicted r^2) also suggest that the limited combined model is best. Adding the full set of chemical factors does not improve the limited combined model. TOC, total fines, and F2 are significant factors in the full model, although total fines has a positive relationship with *H. azteca* biomass. F2 and total fines have the largest coefficients; the coefficient (absolute) for TOC is approxiately 2 times less than those for F2 or total fines.
	habitat-only	4	-36	-25	0.116	5.2	0.0879	yes		8	9	
	limited combined	8	-57	-34	0.277	4.3	0.236	yes	yes	7	8	
	limited chemistry	6	-34	-17	0.116	5.2	0.0773	yes		7	7	
	full	10	-56	-28	0.285	4.4	0.228	no		8	6	
A. abdita survival	null	2	33	38	--	8.1	-0.0199	--		10	72	All models are based on a small data subset (n = 102, down from n = 205). Diagnostic plots indicate that the linear models of *A. abdita* survival are questionable; there are substantial deviations from residual normality and many extreme residual values. Transformation of survival (using log or square root) does not address residual issues. Model statistics suggest that the full model is best; it has the highest adjusted r^2 and predicted r^2 while still having the lowest AIC, suggesting that the model is not overfit. Interestingly, neither the habitat nor limited chemistry variables (F1 through F4) significantly reduce model error relative to the null model, but the combination of all variables results in a significant improvement; adding the entire set of chemical factors improves the model further. Total fines, F2, and F7 are significant factors in the full model (t-test, $p < 0.05$), although total fines is positively related to *A. abdita* survival. F7 is essentially uninterpretable. The coefficients for F2 and total fines are similar, but the coefficient for F7 is less than one-half that for F2.
	habitat-only	4	37	48	-0.0200	8.3	-0.0475	no		6	1	
	limited combined	8	22	43	0.148	7.2	0.0944	yes	yes	5	2	
	limited chemistry	6	34	50	0.0300	8.0	-0.00899	no		3	0	
	full	11	16	45	0.218	6.8	0.149	yes		4	3	

Note: Descriptions of the statistics presented in Table 4-4 are provided in Section 4.1.2.3 MLR, Model Comparison. Shaded rows indicate the best model based on multiple statistics and best professional judgment.

[a] Two separate, sequential F-tests were run by adding a set of models into an ANOVA; these tests compared each model to the model added previously to the ANOVA to determine if there was a significant reduction in model error. The null model was not tested because it was always the first model entered into the ANOVA (with no prior model for comparison). The sequence of models added to the first ANOVA was null, habitat-only, limited combined, and full model, so that the test results indicated the effect of adding habitat variables before chemistry (i.e., whether chemistry improved the model significantly when accounting for habitat effects). The sequence of models added to the second ANOVA was null, limited chemistry, and limited combined models, so that the test indicated the effect of adding habitat variables after chemistry (i.e., whether habitat improved the model significantly when accounting for chemistry effects); the full model was not analyzed in the second ANOVA because the comparison to the limited combined model was the same as for the first ANOVA. In Table 4-4, results from the first ANOVA are reported for the habitat-only, limited combined (left-hand value), and full model; results from the second ANOVA are reported for the limited combined (right-hand value) and limited chemistry models.

ANOVA – analysis of variance
AIC – Akiake's information criterion
BIC – Bayesian information criterion
df – degrees of freedom
HBI – Hilsenhoff Biotic Index
MLR – multiple linear regression
n – sample size
p – number of explanatory variables
PRESS – predicted residual error sum of squares
VIF – variance inflation factor
TOC – total organic carbon
SDI – Swartz's Dominance Index

4.2.2.2 Method 2 Results

Principal Component Analysis

Eigenvalues from the Method 2 PCA indicated that four PCs were required to explain 90% of the total variance in the Method 2 sediment chemistry dataset, and six PCs were required to explain 95% of the total variance.

Exploratory Factor Analysis

Loading values indicated that the factors were associated with:

- F1 – total PCBs and organochlorine pesticides
- F2 – metals, total PAHs, and bis-(2-ethylhexyl) phthalate
- F3 – bis-(2-ethylhexyl) phthalate
- F4 – mercury
- F5 – total PCBs

No chemicals were strongly associated with F6 (Appendix B, Table B2-4). The scores from the Method 2 EFA are provided in Appendix B, Table B2-4.

Multiple Linear Regression

The results of MLR analyses are presented in Table 4-5 (and Appendix B, Table B2-6), which is analogous to the summary of Method 1 results presented in Table 4-4.

Table 4-5. MLR results, Method 2

Endpoint	Model Type	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant ($p < 0.5$)?		No. samples with Cook's D > 4/n	No. Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
Abundance (per m^2)	null	2	222	228	--	43.6	-0.0167	--		4	68	The habitat-only model is best due to its relatively low AIC BIC, and PRESS statistics and relatively high predicted r^2. The addition of chemistry factors does not significantly improve the habitat-only model (i.e., limited combined model); the addition of F5 and F6 significantly improve the model, but these factors are associated with only a small portion of the total variance in sediment chemistry. F5 is mostly associated with PCBs, but PCBs are much more strongly associated with F1, which is not a significant factor in the full model. Based on the parsimony of the habitat-only model and its similarity in fit to the full model, the habitat-only model is best. Regardless, due to the poor fit of all models, this result is highly uncertain. Total fines, F5, and F6 are significant factors in the full model (t-test, $p < 0.05$), although F6 is essentially uninterpretable. The coefficients for significant variables are similar (within a factor of 2).
	habitat-only	4	205	216	0.147	37.6	0.125	yes		1	6	
	limited combined	8	205	227	0.174	37.9	0.117	no	yes	3	0	
	limited chemistry	6	211	227	0.120	40.3	0.0613	yes		4	1	
	full	10	200	228	0.215	35.8	0.165	yes		3	4	
Richness	null	2	803	809	--	5,330	-0.0167	--		6	52	The addition of habitat variables significantly improves the models. The full model is best based on several statistics, including the AIC, PRESS, and r^2 values (adjusted and predicted), suggesting that the full model is not overfit. In the full model, total fines, F1, and F5 are significant factors (t-test, $p < 0.05$). F1 has the highest coefficient among the variables, although the coefficient for total fines is similar. The coefficient for F5 is approximately 3 times less than that for F1.
	habitat-only	4	780	791	0.190	4,424	0.156	yes		5	7	
	limited combined	8	736	759	0.453	3,413	0.349	yes	yes	8	6	
	limited chemistry	6	740	757	0.426	3,541	0.325	yes		9	5	
	full	10	731	759	0.483	3,271	0.376	yes		8	7	
Shannon-Wiener H'	null	2	185	191	--	32.2	-0.0167	--		7	54	The addition of habitat variables does not improve the model that appears to be best, the limited chemistry model. This is based partly on the rule of parsimony (e.g., having 6 variables rather than 8 or 10) and partly on reasonably good fit. The full and limited combined models, while having better fit statistics, do not include any significant variables (t-test, $p < 0.05$) that are not also included in the limited chemistry model (i.e,. F1).
	habitat-only	4	168	179	0.149	28.2	0.11	yes		8	10	
	limited combined	8	115	138	0.465	19.6	0.383	yes	no	7	4	
	limited chemistry	6	116	133	0.453	19.8	0.377	yes		7	5	
	full	10	116	143	0.473	19.7	0.379	no		9	6	
Pielou's J'	null	2	-131	-125	--	2.4	-0.0167	--		6	63	The limited chemistry model is best based on several model statistics (including AIC, BIC, PRESS, predicted r^2, and number of influential points), although the relevance of this result is uncertain due to poor model fit. The addition of habitat variables does not improve the model. F1 is the only significant variable in the full model (t-test, $p < 0.05$).
	habitat-only	4	-132	-121	0.0240	2.4	-0.0114	no		5	6	
	limited combined	8	-139	-117	0.111	2.2	0.0677	yes	no	7	3	
	limited chemistry	6	-141	-125	0.112	2.1	0.0824	yes		5	2	
	full	10	-135	-107	0.097	2.3	0.0245	no		6	3	
SDI (log)	null	2	-34	-29	--	5.3	-0.0167	--		9	50	Q-Q and residual-fitted value plots suggest that there are trends and non-normality in residuals. Log-transformation addresses these issues. The limited chemistry model is best based on several statistics (AIC, BIC, PRESS, and predicted r^2). The inclusion of habitat variables and additional chemical factors does not significantly improve the models. F1 is the only factor significantly related to SDI in the full model (t-test, $p < 0.05$).
	habitat-only	4	-47	-35	0.111	4.8	0.0715	yes		4	8	
	limited combined	8	-91	-68	0.402	3.5	0.313	yes	no	5	4	
	limited chemistry	6	-92	-75	0.399	3.5	0.323	yes		7	5	
	full	10	-91	-63	0.411	3.6	0.303	no		7	6	

Endpoint	Model Type	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant (p < 0.5)?		No. samples with Cook's D > 4/n	No. Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
HBI	null	2	147	152	--	32.1	-0.0288	--		3	40	Residual-fitted value plots indicate a trend in residuals suggesting non-linearity, particularly for the habitat-only model. Basic transformations (i.e., log and square root) do not address these trends. Models are based on a small data subset (n = 71, reduced from n = 121). The limited chemistry model is best based on most model statistics (i.e., AIC, BIC, r^2 values, and PRESS). The addition of habitat variables does not improve the models, and only F1, F2, and F4 are significantly related to HBI (based on the full model; t-test, p < 0.05). F1, F2, and F4 have similar coefficients, all within a factor of 2.
	habitat-only	4	127	136	0.266	24.4	0.219	yes		5	2	
	limited combined	8	95	113	0.555	19.2	0.385	yes	no	6	1	
	limited chemistry	6	92	106	0.564	17.9	0.426	yes		7	1	
	full	10	99	122	0.541	21.1	0.324	no		7	1	
C. dilutus survival	null	2	-85	-80	--	2.2	-0.0216	--		4	59	Models are based on a data subset (n = 94, reduced from n = 121). The full model is best due to its improvement after adding the full list of chemical factors (rather than the limited subset). And, because the PRESS is low and predicted r^2 is high for the full model, it is not likely overfit. Adding habitat variables does not significantly improve the model, so the full model is perhaps not as parsimonious as it could be. Regardless, chemistry factors appear to be more important than either TOC or total fines for predicting *C. dilutus* survival. Due to the weakness of the model fit, this result is highly uncertain. F1, F2, F5, and F6 are significantly related to *C. dilutus* survival (based on the full model; t-test, p < 0.05); F6 is essentially uninterpretable. The coefficients for significant factors are all fairly similar. Due to poor model fits, these results are uncertain.
	habitat-only	4	-86	-76	0.0310	2.3	-0.0856	yes		5	2	
	limited combined	8	-100	-80	0.200	2.2	-0.0179	yes	no	7	1	
	limited chemistry	6	-103	-87	0.206	2.1	0.0404	yes		8	1	
	full	10	-112	-87	0.309	1.9	0.125	yes		7	0	
C. dilutus biomass	null	2	-25	-20	--	4.1	-0.0216	--		4	52	Models are based on a data subset (n = 94, reduced from n = 121). The full model is best due to its improvement model after adding the full list of chemical factors (rather than the limited subset). And, because the PRESS is low and predicted r^2 is high for the full model, it is not likely overfit. Adding habitat variables does not significantly improve the model, so the full model is perhaps not as parsimonious as it could be. Regardless, chemistry factors (i.e., F2, F4, and F5) appear to be more important than either TOC or total fines for predicting *C. dilutus* biomass (based on the full model; t-test, p < 0.05); F2 has the highest coefficient, and the coefficients of F4 and F5 are similar, about half that of F2. Due to the weakness of the model fit, this result is highly uncertain.
	habitat-only	4	-26	-16	0.0300	4.2	-0.0301	no		5	4	
	limited combined	8	-34	-13	0.141	3.9	0.0326	yes	no	5	3	
	limited chemistry	6	-33	-18	0.120	3.9	0.0467	yes		5	4	
	full	10	-42	-16	0.227	3.5	0.123	yes		7	3	
H. azteca survival	null	2	-12	-7	--	6.3	-0.0167	--		9	68	Q-Q plots suggest that residuals are not normally distributed. Transforming survival data does not address these residual issues. The limited combined model is the best, based on several statistics (i.e., AIC, BIC, and PRESS) as well as parsimony (8 parameters vs. 10 in the full model). The addition of the full set of chemical factors does not improve the model, but habitat factors are important. TOC, total fines, F2, and F3 are significantly related to *H. azteca* survival (based on the full model; t-test, p < 0.05), although total fines is positively related to survival. Total fines, F2, and F3 have the largest (similar) coefficients; the coefficient (absolute) for TOC is approximately 3 times less than those for the other variables.
	habitat-only	4	-30	-19	0.151	5.5	0.115	yes		9	4	
	limited combined	8	-41	-19	0.248	5.0	0.187	yes	yes	11	4	
	limited chemistry	6	-34	-17	0.189	5.3	0.145	yes		7	2	
	full	10	-41	-13	0.261	5.0	0.198	no		11	5	
H. azteca biomass	null	2	-23	-18	--	5.8	-0.0167	--		7	59	The limited combined model is best based on all model statistics, and because the inclusion of habitat variables and the limited subset of chemical factors significantly improves the model (although the full set of chemical factors did not improve the model further). Due to the poor fit of all models, this result is highly uncertain. TOC, total fines, F2, F3, and F4 are significantly related to *H. azteca* survival (based on the full model; t-test, p < 0.05), although total fines is positively related to biomass. Total fines, F2, and F3 have the largest coefficients (similar); the coefficients for TOC and F4 are approximately 2 to 3 times less than those for other variables. Due to poor model fits, these results are uncertain.
	habitat-only	4	-36	-25	0.116	5.2	0.0879	yes		8	9	
	limited combined	8	-55	-33	0.269	4.4	0.227	yes	yes	7	5	
	limited chemistry	6	-42	-25	0.168	4.9	0.135	yes		7	7	
	full	10	-53	-25	0.267	4.4	0.215	no		6	6	

Endpoint	Model Type	df	AIC	BIC	Adjusted r^2	PRESS	Predicted r^2	F-test Significant ($p < 0.5$)?		No. samples with Cook's D > 4/n	No. Samples with DFFITS > 2 × sqrt(p/n)	Notes and Model Selection Rationale
A. abdita survival	null	2	-4	-1	--	1.3	-0.0784	--		2	20	Models are based on a data subset (n = 27 [only LPRSA locations], reduced from n = 121). Diagnostic plots indicate that residuals are heteroskedastic, and that extreme values likely influence the regressions. Log-transformation does not address these issues. F6 is collinear with other variables (VIF > 5), so it is excluded from the full model. F6 is essentially uninterpretable chemistry information, so it does not contribute to an interpretable model. Based on ANOVA, no model is a significant improvement over the null model. Therefore, the null model is best: *A. abdita* survival is not related to habitat variables or chemical factors (included in the analysis herein). In the full model, only total fines is significantly related to survival, and that relationship is positive.
	habitat-only	4	-4	2	0.0680	1.4	-0.145	no		3	3	
	limited combined	8	-6	5	0.229	3.2	-1.67	no	yes	5	4	
	limited chemistry	6	1	9	-0.0490	2.6	-1.17	no		3	0	
	full	9	-4	8	0.201	4.0	-2.32	no		6	4	

Note: descriptions of the statistics presented in Table 4-5 are provided in Section 4.1.2.3 “MLR Model Comparison”. Shaded rows indicate the “best” model based on multiple statistics and best professional judgment.

[a] Two separate, sequential F-tests were run by adding a set of models into an ANOVA; these tests compared each model to the model added previously to the ANOVA to determine if there was a significant reduction in model error. The null model was not tested because it was always the first model entered into the ANOVA (with no prior model for comparison). The sequence of models added to the first ANOVA was null, habitat-only, limited combined, and full model, so that the test results indicated the effect of adding habitat variables before chemistry (i.e., whether chemistry improved the model significantly when accounting for habitat effects). The sequence of models added to the second ANOVA was null, limited chemistry, and limited combined models, so that the test indicated the effect of adding habitat variables after chemistry (i.e., whether habitat improved the model significantly when accounting for chemistry effects); the full model was not analyzed in the second ANOVA because the comparison to the limited combined model was the same as for the first ANOVA. In Table 4-4, results from the first ANOVA are reported for the habitat-only, limited combined (left-hand value), and full model; results from the second ANOVA are reported for the limited combined (right-hand value) and limited chemistry models.

ANOVA – analysis of variance
AIC – Akiake’s information criterion
BIC – Bayesian information criterion
df – degrees of freedom
HBI – Hilsenhoff Biotic Index
LPRSA – Lower Passaic River Study Area
MLR – multiple linear regression
n – sample size
p – number of explanatory variables
PCB – polychlorinated biphenyl
PRESS – predicted residual error sum of squares
VIF – variance inflation factor
TOC – total organic carbon
SDI – Swartz’s Dominance Index

4.2.2.3 Discussion of Method 1 and Method 2 results

The results of Method 1 and Method 2 analyses were generally consistent across benthic invertebrate community metric and sediment toxicity test endpoints, in that similar model types were selected (among the five types described in Section 4.1.2.3). In general, the MLR models had poor predictive power, as evidenced by several statistics (Tables 4-4 and 4-5). Sediment chemistry was nearly always an important factor in the selected MLR models, with few exceptions (i.e., abundance and *A. abdita* survival in Table 4-5).

Of the response variables modeled, sediment chemical factors and total fine-grained sediment appeared to be the dominant explanatory variable(s) (in terms of magnitude and significance of coefficients) in many full MLRs. Chemical factors were significant in models for richness, Shannon-Wiener H', Pielou's J', HBI, *C. dilutus* survival and biomass, and *H. azteca* survival and biomass. Habitat variables were also significant in many of those models, although in some cases, they had coefficients of the opposite sign (suggesting that habitat variables may ameliorate stress).

Based on all F-test results comparing the relative importances of habitat variables and sediment chemical factors (Tables 4-4 and 4-5), sediment chemical factors were more important than habitat variables[51] in one-half of model evaluations (11 of 22), and were equally as important as habitat variables in nearly one-half of the remaining model evaluations (9 of 22). In the case of abundance (Table 4-5), habitat variables were more important than chemical factors, and in the case of *A. abdita* survival (Table 4-5), neither factor improved the model fit relative to the null model.

The results from the modeling of abundance were conflicting; based on Method 1 and Method 2, the full model and habitat-only model were selected, respectively. The relative importance of habitat variables and chemical factors switched between the two methods, with chemical factors being marginally more important in the full model of Method 1, and habitat variables being more important in the habitat-only model of Method 2. Due to these conflicting results, the results from both methods were uncertain. Abundance may either increase or decrease in response to environmental stress, so the lack of a clear and consistent linear relationship with chemical or habitat variables is not unreasonable.

A. abdita survival appeared to be very weakly related or unrelated to sediment chemical factors or habitat variables based on the analysis. In Method 1, the full model was selected because a significant improvement over the null model required that all variables be included. In Method 2, the null model was selected because, even with all chemical and habitat variables, a significant model was not developed. Small sample

[51] Sediment chemical factors were more important than habitat variables when adding habitat variables to the MLR did not improve model fit after controlling for chemical factors, but adding chemical factors did improve the model fit after controlling for habitat variables (in a separate F-test).

size may have played a part in the lack of observed significance in the Method 2 analysis; only 27 samples were available to develop *A. abdita* survival models.

Models developed for HBI tended to be the strongest, with adusted r^2 values of 0.678 and 0.564 for Method 1 and Method 2, respectively. These models indicated that chemical factors were more important for predicting HBI than habitat variables. One uncertainty associated with these models in particular was the comparatively small datasets on which they were based (i.e., freshwater samples only). Still, the sample sizes for HBI model development were 72 and 71 for Method 1 and Method 2, respectively, which should have been sufficient to observe significant trends.

4.2.3 T20 and T50 threshold comparison

LPRSA locations with sediment chemistry exceeding one or more T20 or T50 are described in Table 4-6. This information is also presented in Appendix B, Table B7-1.

Table 4-6. T20 and T50 screen of LPRSA sediment chemistry data

Location ID	No. of Chemicals Exceeding T20	No. of Chemicals Exceeding T50
LPRT01A	35	26
LPRT01B	36	25
LPRT01C	35	26
LPRT01D	37	29
LPRT01E	35	24
LPRT01F	36	30
LPRT01G	37	28
LPRT02A	37	27
LPRT02B	37	28
LPRT02C	36	30
LPRT02D	36	30
LPRT02E	36	32
LPRT02F	37	31
LPRT03A	37	31
LPRT03B	36	30
LPRT03C	36	31
LPRT03D	37	32
LPRT03E	37	31
LPRT03F	36	31
LPRT04A	24	3
LPRT04B	37	31
LPRT04C	37	30
LPRT04D	36	32

Table 4-6. T20 and T50 screen of LPRSA sediment chemistry data

Location ID	No. of Chemicals Exceeding T20	No. of Chemicals Exceeding T50
LPRT04E	37	30
LPRT04F	36	28
LPRT05A	37	31
LPRT05B	37	31
LPRT05C	35	31
LPRT05D	37	32
LPRT05E	36	28
LPRT05F	36	30
LPRT06A	37	30
LPRT06B	37	31
LPRT06C	36	31
LPRT06D	36	30
LPRT06E	36	30
LPRT06F	36	31
LPRT07A	33	27
LPRT07B	36	31
LPRT07C	37	33
LPRT07D	37	31
LPRT07E	37	33
LPRT08A	37	30
LPRT08B	36	31
LPRT08C	37	34
LPRT08D	35	25
LPRT08E	30	13
LPRT09A	36	27
LPRT09B	31	22
LPRT09C	33	19
LPRT09D	36	31
LPRT09E	36	30
LPRT09F	36	31
LPRT09G	27	1
LPRT09H	27	6
LPRT10A	36	31
LPRT10B	36	31
LPRT10C	35	30

Table 4-6. T20 and T50 screen of LPRSA sediment chemistry data

Location ID	No. of Chemicals Exceeding T20	No. of Chemicals Exceeding T50
LPRT10D	28	15
LPRT10E	36	32
LPRT11A	30	20
LPRT11B	36	31
LPRT11C	33	20
LPRT11D	34	29
LPRT11E	36	31
LPRT11F	35	27
LPRT11G	37	34
LPRT12A	36	31
LPRT12B	36	31
LPRT12C	36	31
LPRT12D	36	23
LPRT12E	37	32
LPRT13A	33	27
LPRT13B	36	29
LPRT13C	36	31
LPRT13D	28	2
LPRT13E	35	28
LPRT13F	33	23
LPRT13G	36	29
LPRT14A	33	22
LPRT14B	33	26
LPRT14C	30	23
LPRT14D	34	24
LPRT14E	31	22
LPRT14F	31	23
LPRT15A	33	23
LPRT15B	36	29
LPRT15C	28	20
LPRT15D	30	17
LPRT15E	30	21
LPRT15F	28	22
LPRT16A	26	9
LPRT16B	30	21

Table 4-6. T20 and T50 screen of LPRSA sediment chemistry data

Location ID	No. of Chemicals Exceeding T20	No. of Chemicals Exceeding T50
LPRT16C	25	11
LPRT16D	27	21
LPRT16E	33	24
LPRT17A	34	27
LPRT17D	31	24

ID – identification
LPRSA – Lower Passaic River Study Area
T20 – 20% probability of observing toxicity
T50 – 50% probability of observing toxicity

As can be seen in Table 4-5, at least one T50 value was exceeded (resulting in the prediction of toxicity) at every LPRSA SQT location. Given that sediment toxicity was not actually observed at all LPRSA SQT locations (Section 3.2.1 or Table 3-6), this result (and its implication for risk characterization in the BERA) is questionable. Discussion of uncertainty associated with the T20 and T50 screen of LPRSA data is provided in Section 4.3.1.

4.3 UNCERTAINTY ANALYSES

The following sections describe analyses of uncertainty associated with the sediment chemistry LOE. Sections 4.3.1 through 4.3.3 describe several quantitative analyses of uncertainty, and Section 4.3.4 provides additional qualitative statements of uncertainty.

4.3.1 Reliability analysis

4.3.1.1 T20 and T50 values

Table 4-7 presents results of the reliability of the T50 values for predicting toxicity in the LPRSA relative to reference conditions. As noted in Section 4.2.4, at least one T50 value was exceeded at all LPRSA locations, so all locations were predicted to be toxic. As a result, there were neither true nor false negatives associated with the reliability analysis (Table 4-7). This also resulted in sensitivity and specificity values of 1 and 0, respectively, which clearly indicated that the T50 was unable to correctly classify LPRSA locations where there was no toxicity relative to reference conditions. Although the T50 correctly predicted toxicity where toxicity was observed in the LPRSA, it did so only by predicting toxic effects at every location. These results highlight the significant uncertainty associated with the T50, which may be an inappropriate (i.e., overly conservative) criterion for categorizing LPRSA locations in the sediment chemistry LOE.

Table 4-7. Reliability of T50 values for predicting toxicity using site-specific LPRSA toxicity

Reliability Statistic[a]	*C. dilutus* Survival	*C. dilutus* Biomass	*H. azteca* Survival	*H. azteca* Biomass	*A. abdita* Survival
Sample size	71	71	98	98	27
True positives	13	58	50	34	15
False positives	58	13	48	64	12
True negatives	0	0	0	0	0
False negatives	0	0	0	0	0
Type I error rate	0%	0%	0%	0%	0%
Type II error rate	82%	18%	49%	65%	44%
Total error rate	82%	18%	49%	65%	44%
Sensitivity	1	1	1	1	1
Specificity	0	0	0	0	0
Precision	0.18	0.82	0.51	0.35	0.56
False discovery rate	0.82	0.18	0.49	0.65	0.44

[a] Reliability statistics are described in the text (Section 4.1.4.1).

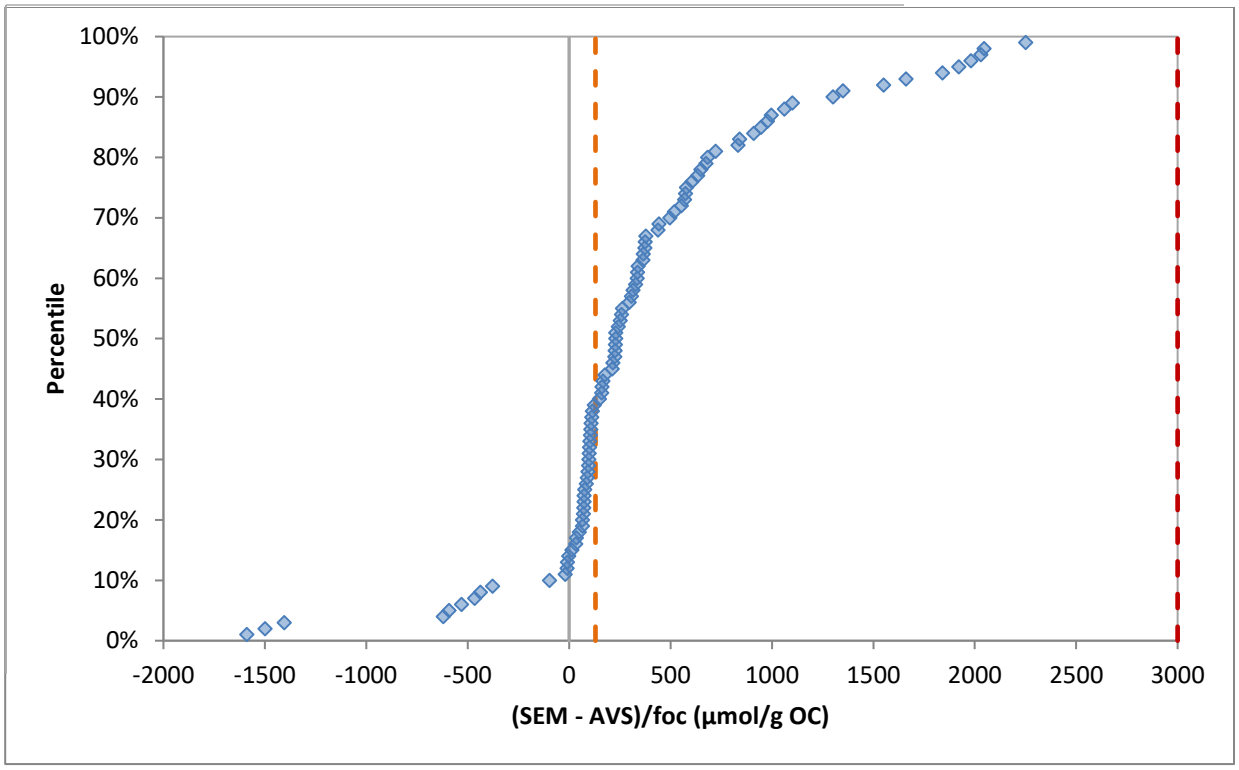
LPRSA – Lower Passaic River Study Area

T50 – 50% probability of observing toxicity

4.3.2 SEM-AVS analysis

In the LPRSA, the OC-normalized SEM-AVS concentrations[52] for SQT sediment samples ranged from -1,590 to 2,250 µmol/g OC, with an average value of 359 µmol/g OC. All of the OC-normalized SEM-AVS concentrations fell below the 3,000 µmol/g OC threshold identified as the threshold above which there is a high likelihood of toxicity (USEPA 2005b) (i.e., no SEM concentrations with high likelihood of toxicity), and 40% (39 of 98) of the SQT samples had OC-normalized SEM-AVS concentrations that were below the 130 µmol/g OC threshold (i.e., no toxicity expected) (Figure 4-1). Of the remaining samples, 60% (59 of 98) had OC-normalized SEM-AVS concentrations that were within the range identified as being uncertain as a predictor of toxicity (USEPA 2005b) (i.e., > 130 µmol/g OC and ≤ 3,000 µmol/g OC). In addition, only 12 of the 59 sediment samples with OC-normalized SEM-AVS concentrations that were within the uncertain level of toxicity had concentrations greater than 1,000 µmol/g OC.

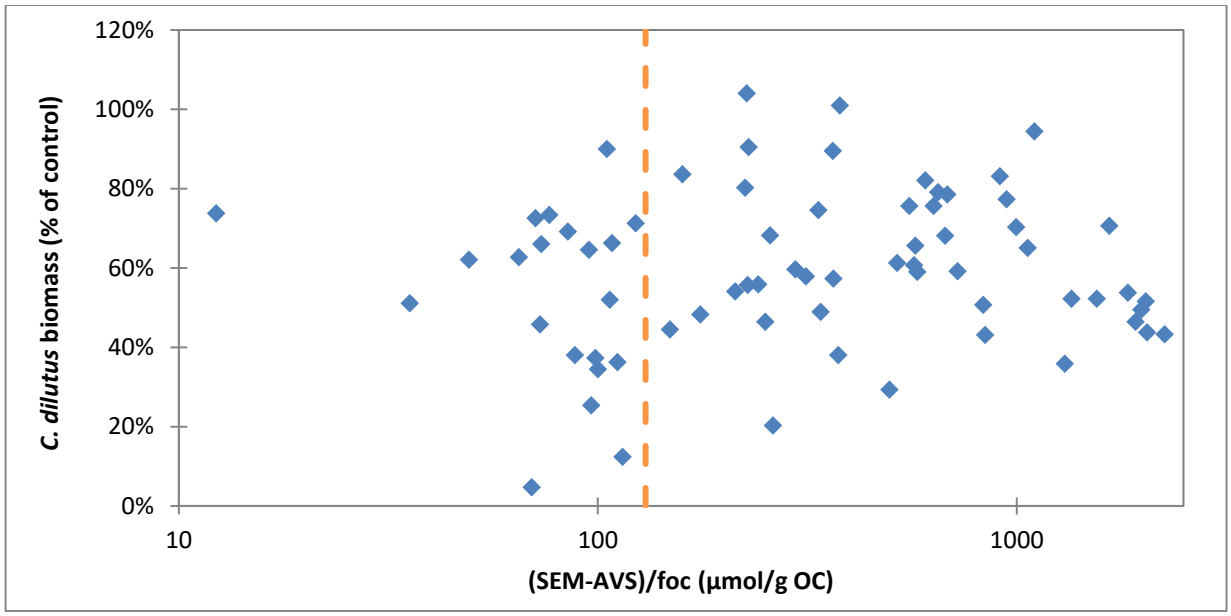
[52]Values are based on data presented in Appendix K. The OC-normalized SEM-AVS concentrations are calculated as: $(\Sigma SEM–AVS)/f_{oc}$; ΣSEM is the sum of SEM (i.e., cadmium, copper, nickel, lead, silver, and zinc), and f_{oc} is the fractional value of OC (reported in Appendix K as percent TOC).



Note: Vertical lines mark toxicity thresholds at 130 µmol/g OC and 3,000 µmol/g OC (USEPA 2005b).

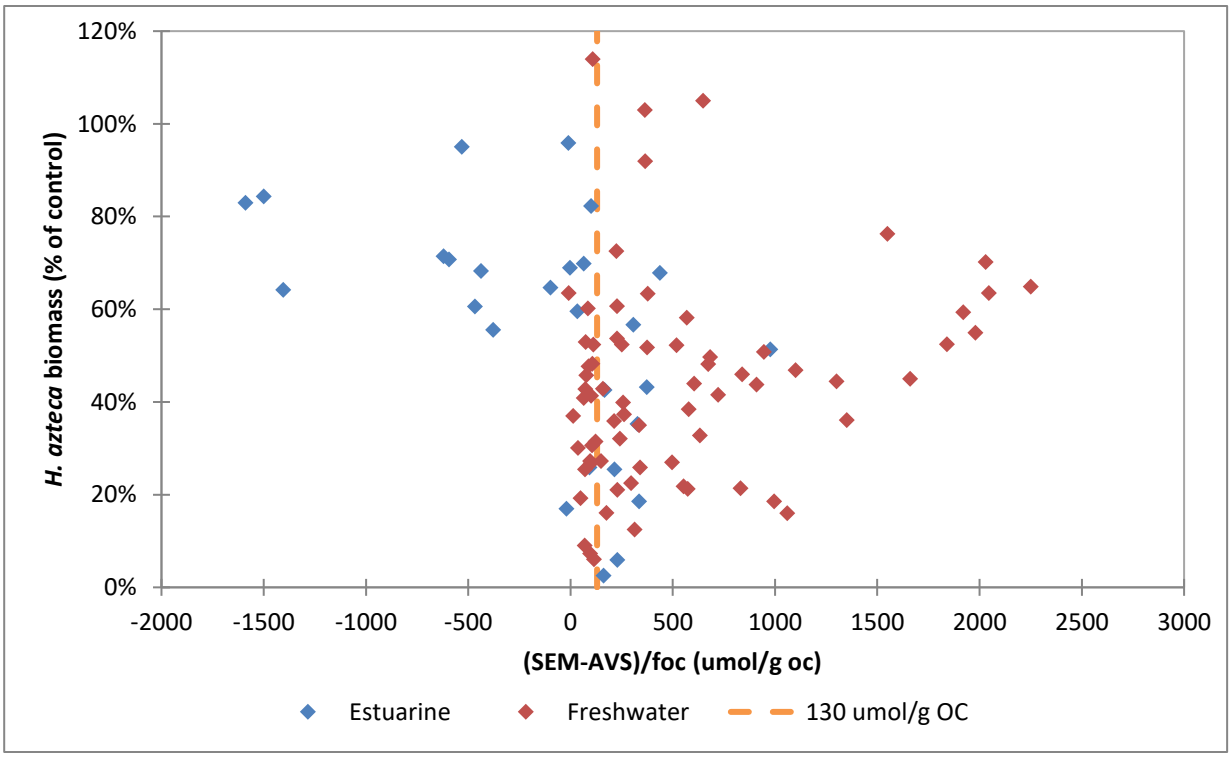
Figure 4-1. Cumulative frequency distribution of SEM-AVS normalized to OC for LPRSA SQT sediment samples

Example plots of LPRSA toxicity response data and OC-normalized SEM-AVS provided in Figures 4-2 through 4-4 do not show a clear relationship between toxic responses and OC-normalized SEM-AVS concentrations measured in LPRSA sediment. Other endpoints (not shown) have similarly unclear or negligible relationships with SEM-AVS.



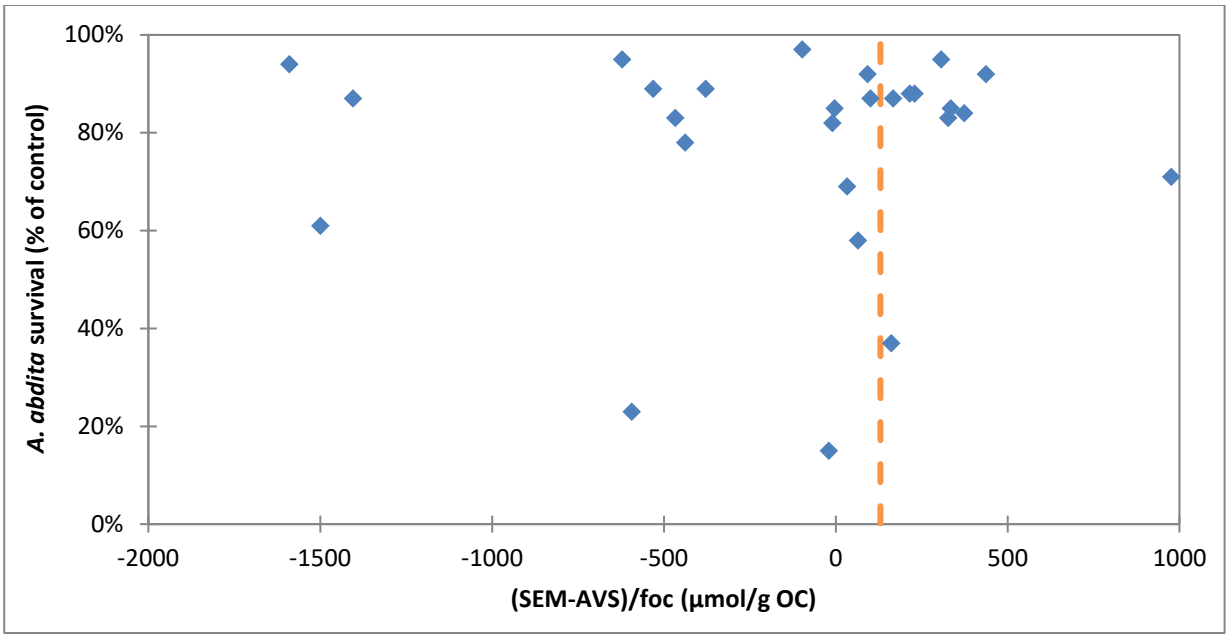
Note: *C. dilutus* biomass data are control normalized and presented as percent of control. Vertical line marks the low toxicity threshold at 130 µmol/g OC (USEPA 2005b); no values exceed the 3,000 µmol/g OC threshold.

Figure 4-2.*C. dilutus* biomass and (SEM-AVS)/f_{OC} in the LPRSA, fall 2009



Note: *H. azteca* biomass data are control normalized and presented as percent of control. Vertical line marks the low toxicity threshold at 130 µmol/g OC (USEPA 2005b); no values exceed the 3,000 µmol/g OC threshold. Different symbols are used to show standard ("freshwater") and salinity-acclimated ("estuarine") *H. azteca* test results.

Figure 4-3.*H. azteca* biomass and (SEM-AVS)/f_{OC} in the LPRSA, fall 2009



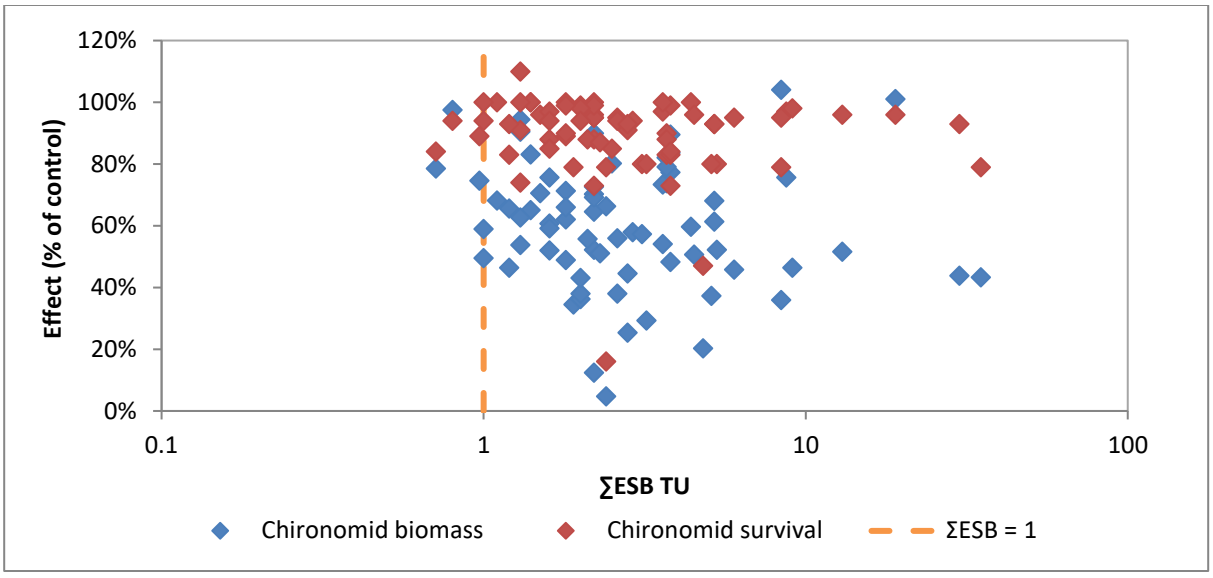
Note: *A. abdita* survival data are control normalized and presented as percent of control. Vertical line marks the low toxicity threshold at 130 µmol/g OC (USEPA 2005b); no values exceed the 3,000 µmol/g OC threshold.

Figure 4-4.*A. abdita* survival and (SEM-AVS)/f_{OC} in the LPRSA, fall 2009

Based on the evaluation of SEM-AVS in LPRSA sediment presented above, OC-normalized SEM-AVS observed in LPRSA samples is not a useful predictor of sediment toxicity and is of limited value in assessing benthic community risk. The only exception appears to be *H. azteca* biomass measured in sediments from estuarine toxicity test locations (interstitial salinity ≥ 5 ppt), but the relationship is inconsistent with that of the same test endpoint measured in freshwater toxicity tests. Therefore, the relationship between *H. azteca* biomass and SEM-AVS remains uncertain.

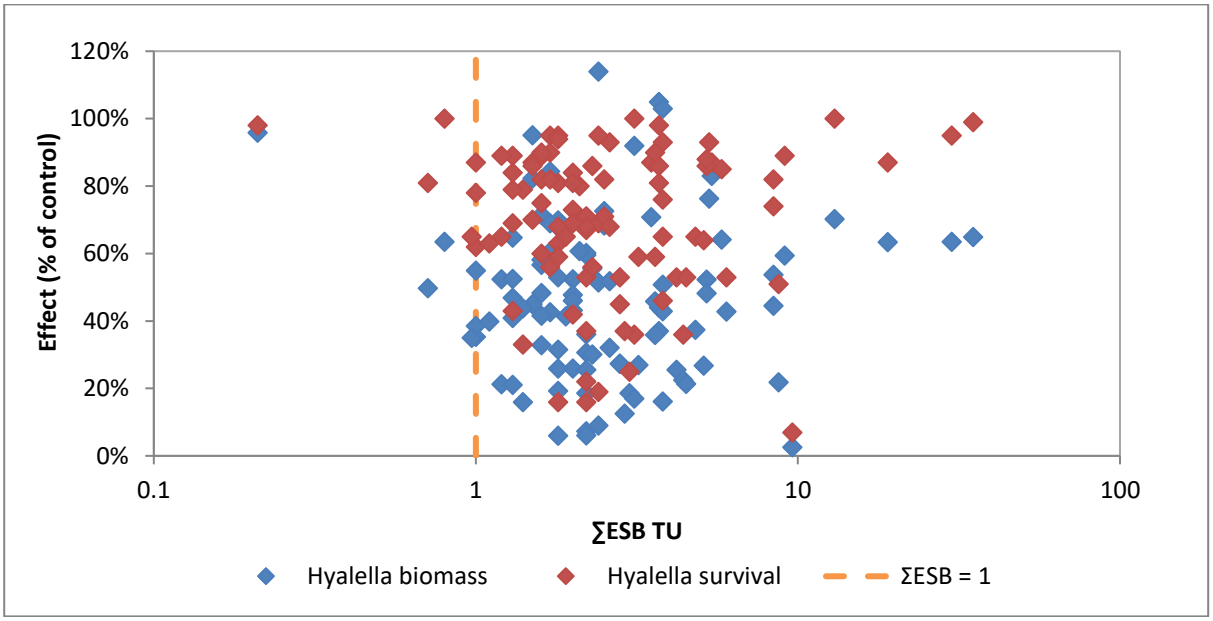
4.3.3 Equilibrium partitioning-based analysis of 34 PAHs

In the LPRSA, the 34-PAH ΣESB TUs for SQT sediment samples ranged from 0.21 to 35, with an average TU of 3.7. The TUs from 91 of 98 LPRSA locations (93%) were greater than 1, which USEPA (2003) suggests may result in toxicity in sensitive aquatic species. Figures 4-5 through 4-7 provide visual comparisons of calculated TU values to sediment toxicity test results. These three figures each show an inconsistent or nonexistent relationship between PAHs and toxicity.



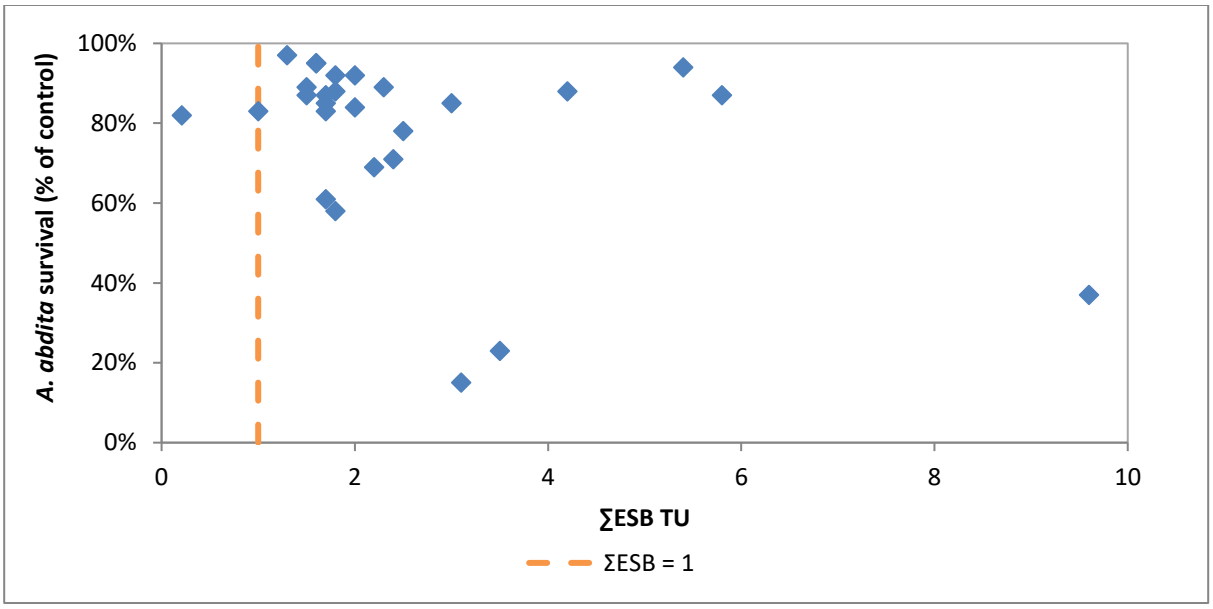
ΣESB TU – sum equilibrium partitioning sediment benchmark toxic units

Figure 4-5.Relationship between *C. dilutus* toxicity and 34-PAH TUs



ΣESB TU – sum equilibrium partitioning sediment benchmark toxic units

Figure 4-6.Relationship between *H. azteca* toxicity and 34-PAH TUs



∑ESB TU – sum equilibrium partitioning sediment benchmark toxic units

Figure 4-7.Relationship between *A. abdita* survival and 34-PAH TUs

Based on the analysis of 34-PAH TUs from LPRSA SQT data, it appears that the EqP approach is not a useful predictor of LPRSA sediment toxicity and is of limited value in assessing benthic community risk.

4.3.4 Additional points of uncertainty

In addition to uncertainties described in Sections 4.3.1 through 4.3.3, the following uncertainties are associated with the sediment chemistry LOE:

- Correlation-based analyses (i.e., Spearman rank correlation) do not determine causality, but causality is assumed to be related to the strength of significant correlations (r-values); in complex systems, multiple correlations may exist, including both independent covariance (e.g., multiple chemicals from a similar source) and dependent covariance (interrelation) (e.g., chemical concentrations and sediment grain size). Interrelated factors can also interact (i.e., additive, antagonistic, or synergistic effects), potentially leading to non-linear and non-monotonic relationships. These complex interactions are not identified using correlation analysis, and they may weaken correlations and result in inaccurate inference.
- Spearman correlation analysis may not be appropriate for benthic response variables that are non-monotonically related to environmental stress (e.g., abundance) (Weisberg et al. 1997); abundance was treated as monotonic in this LOE. This uncertainty has been addressed in the benthic invertebrate community LOE (Section 2).

- The interpretation of PCs and factors (derived using PCA and EFA, respectively) is somewhat subjective. The assignment of groups or single chemicals to PCs or factors according to their loading values introduces a level of inherent error into the interpretation of PCA or EFA results because PCs and factors only explain a portion of the variance in the original variables. By focusing on only the first few PCs or factors from each PCA or EFA (i.e., those that explain the greatest amount of the overall variance in the datasets), the uncertainty associated with the interpretation of PCs or factors (in terms of the original chemical variables) is minimized to a reasonable extent.
- Although MLR model fits were significant for all but one benthic invertebrate response variable (i.e., *A. abdita* survival based on Method 2), the ability of the MLR models to predict effects is questionable. Model statistics (Tables 4-4 and 4-5) indicate poor fits for nearly all models. This may be due to several factors, including weak relationships between chemical factors, habitat variables, and benthic response variables or relationships that exist but are nonlinear. Nonlinear relationships were not evaluated as part of the multivariate assessment of SQT data.
- SEM-AVS is difficult to measure precisely, and measurements are irreproducible among laboratories (Hammerschmidt and Burton 2010). LPRSA sediment analyses were conducted by a single laboratory, thereby reducing the uncertainty associated with interlaboratory error, but the accuracy of measurements remains uncertain.
- Using T20 and T50 values for the sediment chemistry LOE is likely inappropriate for defining ecological risks in a BERA. Field et al. (2002) note that the "LRM approach provides a useful framework for conducting screening-level assessments…" and that the model does not consider site-specific bioavailability or exposure. Furthermore, the T20 and T50 values are based on field-collected sediments rather than controlled sediments (Field et al. 2002), so they are likely to contain more hazardous substances than the one for which the criteria were developed. For example, a T50 value for total PAHs may be based on sediment that contains PAHs, PCBs, metals, and any number of other stressors. As a result, it is likely that T20 and T50 values may overestimate the toxicity of the single contaminants for which they are reported. The quantitative analysis of uncertainty incorporates reference area toxicity data to establish a study-specific mean-quotient threshold, rather than applying generic, non-site-specific criteria.
- Mean-quotients, like T20s and T50s, are based on criteria primarily meant for screening purposes and that are based (at least in part) on data from field-collected sediment (with mixtures of sediment contaminants) (Long et al. 1995; MacDonald et al. 2000; Wenning et al. 2005).

- The logistic modeling approach used to derive T20 and T50 values does not address the magnitude of the relationship between concentration and "toxic" response. The T20 and T50 values correspond to 20 and 50% probabilities, respectively of observing > 10% reduction in amphipod survival and a significant difference from the negative control results. Therefore, it is not possible to determine what level of effect can be expected (i.e., what magnitude of risk to invertebrates) from exceedances of T20 and T50 values. The use of mean-quotient values as part of the quantitative analysis of uncertainty addresses the magnitude of possible effects by scaling sediment concentrations to those correlated in the literature with toxic impacts.
- A high mERMq threshold could not be developed for non-urban reference conditions because acceptable *A. abdita* survival was observed even at the highest calculated mERMq. The quantitative analysis of uncertainty for the sediment chemistry LOE, comparing LPRSA conditions to non-urban reference conditions, was based on mERMq thresholds developed using the urban reference condition. This resulted in a more conservative application of sediment chemistry LOE weights for the quantitative analysis of uncertainty based on the comparison of LPRSA data to non-urban reference conditions.
- Sediment concentrations were screened in Appendix A to the BERA to identify COPECs relevant to benthic invertebrates, but those COPECs were not directly addressed in the sediment chemistry LOE described herein. Table 4-8 provides a summary of the sediment chemistry COPECs for benthic invertebrates determined in the SLERA (Appendix A); Table 4-8 also details which COPECs were addressed in part through the mean-quotient, T20/T50 screening analyses or bivariate and multivariate analyses. Due to limited availability of screening-level sediment guidelines, not all COIs from LPRSA sediments were screened in the SLERA or analyzed further in this LOE. It is possible, though unlikely, that benthic invertebrate risks in the LPRSA are caused (at least in part) by COIs that were not evaluated.

Table 4-8. Overlap between surface sediment COPECs for benthic invertebrates and sediment chemistry LOE analyses

COPEC Determined in SLERA	Evaluated Using Bivariate and Multivariate Analyses?[a]	Criterion Available?[a]			
		ERM	PEC	T20	T50
Metals					
Antimony	X			X	X
Arsenic	X	X	X	X	X
Cadmium	X	X	X	X	X
Chromium	X	X	X	X	X
Cobalt[b]					

Table 4-8. Overlap between surface sediment COPECs for benthic invertebrates and sediment chemistry LOE analyses

COPEC Determined in SLERA	Evaluated Using Bivariate and Multivariate Analyses?[a]	Criterion Available?[a]			
		ERM	PEC	T20	T50
Copper	X	X	X	X	X
Lead	X	X	X	X	X
Mercury	X	X	X	X	X
Methylmercury[b]					
Nickel	X	X	X	X	X
Selenium[b]					
Silver	X	X		X	X
Vanadium[b]					
Zinc	X	X	X	X	X
Butyltin					
TBT[b]					
PAHs					
1-methylnaphthalene	X			X	X
1-methylphenanthrene	X			X	X
2,6-dimethylnaphthalene				X	X
2-methylnaphthalene	X	X		X	X
Acenaphthene	X	X		X	X
Acenaphthylene	X	X		X	X
Anthracene	X	X	X	X	X
Benzo(a)anthracene	X	X	X	X	X
Benzo(a)pyrene	X	X	X	X	X
Benzo(b/j)fluoranthene				X	X
Benzo(e)pyrene[b]					
Benzo(g,h,i)perylene	X			X	X
Benzo(k)fluoranthene				X	X
Chrysene	X	X	X	X	X
Dibenzo(a,h)anthracene	X	X		X	X
Fluoranthene	X	X	X	X	X
Fluorene	X	X	X	X	X
Indeno(1,2,3-cd)pyrene	X			X	X
Naphthalene	X	X	X	X	X
Perylene				X	X

Table 4-8. Overlap between surface sediment COPECs for benthic invertebrates and sediment chemistry LOE analyses

COPEC Determined in SLERA	Evaluated Using Bivariate and Multivariate Analyses?[a]	Criterion Available?[a]			
		ERM	PEC	T20	T50
Phenanthrene		X	X	X	X
Pyrene	X	X	X	X	X
Total benzofluoranthenes[b]					
Total HPAHs					
Total LPAHs					
Total PAHs	X	X	X		
SVOCs					
2,4-Dinitrotoluene[b]					
2,6-Dinitrotoluene[b]					
4-Methylphenol[b]					
Bis-(2-ethylhexyl)phthalate	X				
Butylbenzylphthalate[b]					
Dibenzofuran[b]					
Diethylphthalate[b]					
Dimethylphthalate[b]					
Di-n-butylphthalate[b]					
Di-n-octylphthalate[b]					
Isophorone[b]					
n-Nitrosodiphenylamine[b]					
Pentachlorophenol[b]					
Phenol					
PCB aroclors					
Aroclor-1254[b]					
Aroclor-1260[b]					
PCB Congeners					
Total PCB Congeners	X	X	X	X	X
PCB TEQ – Fish[b]					
PCDDs/PCDFs					
2,3,7,8-TCDD[b]					
PCDD/PCDF TEQ - Fish[b]					
Total TEQ-Fish[b]					

Table 4-8. Overlap between surface sediment COPECs for benthic invertebrates and sediment chemistry LOE analyses

COPEC Determined in SLERA	Evaluated Using Bivariate and Multivariate Analyses?[a]	Criterion Available?[a]			
		ERM	PEC	T20	T50
Organochlorine pesticides					
4,4'-DDD				X	X
4,4'-DDE		X		X	X
4,4'-DDT				X	X
Aldrin[b]					
alpha-BHC[b]					
alpha-Chlordane[b]	X				
beta-BHC[b]					
Dieldrin	X		X	X	X
Endosulfan I[b]					
Endosulfan II[b]					
Endrin			X		
gamma-BHC (Lindane)			X		
gamma-Chlordane[b]					
Heptachlor[b]					
Heptachlor epoxide			X		
Hexachlorobenzene[b]	X				
Methoxychlor[b]					
Total chlordane			X		
Total DDx	X	X	X		
Herbicides					
2,4,5-TP (Silvex) [b]					
VOCs					
1,2,3-Trichlorobenzene[b]					
1,2,4-Trichlorobenzene[b]					
1,4-Dichlorobenzene[b]					
1,4-Dioxane[b]					
Acetone[b]					
m, p-Xylene[b]					
Toluene[b]					
Wet chemistry					

Table 4-8. Overlap between surface sediment COPECs for benthic invertebrates and sediment chemistry LOE analyses

COPEC Determined in SLERA	Evaluated Using Bivariate and Multivariate Analyses?[a]	Criterion Available?[a]			
		ERM	PEC	T20	T50
Cyanide[b]					

Note: Surface sediment COPECs for benthic invertebrates were determined in the SLERA (Appendix A).

[a] Values of "X" indicate either that a COPEC was evaluated using the bivariate or multivariate statistical analyses, or that the specified criterion is available for the given COPEC.

[b] COPEC was not directly addressed in the sediment chemistry LOE. COPECs may have been indirectly addressed as part of sums; for example benzo[a]pyrene is a component of both total HPAHs and total PAHs.

BHC – benzene hexachloride
COPEC – chemical of potential ecological concern
DDD – dichlorodiphenyldichloroethane
DDE – dichlorodiphenyldichloroethylene
DDT – dichlorodiphenyltrichloroethane
ERM – effect range-median
HPAH – high-molecular-weight polycyclic aromatic hydrocarbon
LOE – line of evidence
LPAH – low-molecular-weight polycyclic aromatic hydrocarbon
SLERA – Screening Level Ecological Risk Assessment
PAH – polycyclic aromatic hydrocarbon
PCB – polychlorinated biphenyl
PCDD – polychlorinated dibenzo-*p*-dioxin
PCDF – polychlorinated dibenzofuran
PEC – probable effects concentration
SLERA – screening-level ecological risk assessment
SVOC – semivolatile organic compound
T20 – 20% probability of observing toxicity
T50 – 50% probability of observing toxicity
TCDD – tetrachlorodibenzo-*p*-dioxin
TEQ – toxic equivalent

4.3.5 Quantitative analysis of uncertainty

The sediment chemistry LOE approach that is being carried forward into the WOE analysis (Section 6.1 of the BERA main text) is based on the screen of LPRSA sediment chemistry data against T20 and T50 values from the literature (Section 4.2.4). To address key uncertainties associated with that approach (Sections 4.3.1 and 4.3.4), a quantitative analysis was conducted (Appendix B, Table B7-2 and B7-3). The results from that approach are described in the following sections.

4.3.5.1 Mean-quotient threshold comparison

Based on reference area sediment toxicity data and calculated mERMq and mPECq values, the following low and high mERMq and mPECq thresholds were established for screening LPRSA sediment chemistry data (i.e., mERMq and mPECq values):

- Low mERMq thresholds:
  - Jamaica Bay: 0.020
  - Mullica River/Great Bay: 0.055
- High mERMq thresholds:
  - Jamaica Bay: 1.8
  - Mullica River/Great Bay: > 0.27

- Low mPECq threshold: < 0.076
- High mPECq threshold: 1.9

The low mPECq threshold could not be accurately determined because both *C. dilutus* and *H. azteca* survival were < 75% of the negative control at the lowest calculated mPECq value from above Dundee Dam (0.076). As a result, freshwater LPRSA locations could not be classified as having no potential for toxicity (i.e., all were given a weight of at least 0.5). Similarly, the high mERMq threshold for Mullica River/Great Bay could not be determined because there was > 75% survival of *A. abdita* (relative to control) at the highest calculated mERMq. In this case, it was not possible to accurately assign a high weight (1.0) to LPRSA data; as a result, the thresholds for Mullica River/Great Bay were not used. Instead, the Jamaica Bay mERMq thresholds were used to evaluate the sediment chemistry LOE for the comparison of LPRSA data to both urban and non-urban reference areas, which resulted in a more conservative analysis.

Of the 98 LPRSA locations where toxicity and sediment chemistry were measured, 72 (73%) were categorized as having unclear potential for sediment toxicity (as defined by the reference area datasets). No station was categorized as having negligible potential for toxicity, and 26 of 98 locations (27%) were categorized as having high potential for toxicity. The full results of the mean-quotient threshold comparison are presented in Appendix B, Tables B7-2 and B7-3.

The result of screening LPRSA sediment chemistry data using the T20 and T50 values (Section 4.2.4) is in stark contrast to the result of using the mean-quotient approach. The latter approach resulted in only 27% of LPRSA SQT locations being classified as having high potential for toxicity, while the approach using the T20 and T50 values resulted in 100%. Uncertainties associated with the mean-quotient approach are presented in Section 4.3.5.2.

4.3.5.2 Mean-quotient threshold reliability analysis

The reliability of mean-quotient thresholds for predicting toxicity in the LPRSA relative to reference conditions was evaluated similarly to the reliability of T50 values (Section 4.3.1.1), (but using mERMq or mPECq for estuarine or freshwater locations, respectively).[53] The results of this evaluation are presented in Table 4-9.

Table 4-9. Reliability of high mean-quotient threshold for predicting toxic impacts using site-specific LPRSA toxicity

Reliability Statistic[a]	*C. dilutus* Survival[b]	*C. dilutus* Biomass[b]	*H. azteca* Survival[b]	*H. azteca* Biomass[b]	*A. abdita* Survival[b]
Sample size	71	71	98	98	27

[53] Freshwater and estuarine locations are defined as those with interstitial water salinities < 5 ppt or ≥ 5 ppt, respectively.

Reliability Statistic[a]	*C. dilutus* Survival[b]	*C. dilutus* Biomass[b]	*H. azteca* Survival[b]	*H. azteca* Biomass[b]	*A. abdita* Survival[b]
True positives	2	7	14	10	10
False positives	6	1	12	16	8
True negatives	52	12	36	48	4
False negatives	11	51	36	24	5
Type I error rate	15%	72%	37%	24%	19%
Type II error rate	8%	1%	12%	16%	30%
Total error rate	24%	73%	49%	41%	48%
Sensitivity	0.15	0.12	0.28	0.29	0.67
Specificity	0.90	0.92	0.75	0.75	0.33
Precision	0.25	0.88	0.54	0.38	0.56
False discovery rate	0.75	0.13	0.46	0.62	0.44

[a] Reliability statistics are described in Section 4.1.4.1.

[b] Toxicity data are control normalized.

LPRSA – Lower Passaic River Study Area

There appears to be a trade-off in conservatism between the use of logistic regression model-based T20s or T50s and the mean-quotient approach. Specifically, T50s are associated with excessive false positives (i.e., 100% of LPRSA locations were predicted to be toxic), and mean-quotients are associated with numerous false negatives, particularly for *C. dilutus* biomass (72% error rate). Prediction error rates for *H. azteca* and *A. abdita* endpoints were more balanced when using the mean-quotient approach, although total error was slightly higher when using the mean-quotient approach to predict *A. abdita* survival (44% compared to 48%).

Sensitivities and specificities also highlight the trade-off in conservatism between T50s and mean-quotients. For example, the sensitivities are 1.0 for all endpoints of T50s (Table 4-7), whereas specificity is very high for the mean-quotient approach (Table 4-9). This indicates that T50s will always predict toxicity correctly when it is actually observed, whereas mean-quotients will generally be able to correctly predict if toxicity will not occur when it is not actually observed. The mean-quotient approach is more reliable than the T50, in that the former can predict both toxicity (i.e., sensitivities range from 0.12 to 0.67) and non-toxicity (i.e., specificities range from 0.33 to 0.92). Overall, the mean-quotient approach is more reliable (although less conservative) than the T50 approach for predicting impacts (relative to reference area conditions) on *C. dilutus* survival, *H. azteca* survival and biomass, and *A. abdita* survival.

4.4 SUMMARY

Based on the various discussions of sediment chemistry presented in Section 4.2 (and subsections) and uncertainties in Section 4.3 (and subsections), sediment chemistry data appear to be significantly related to observable effects on benthic invertebrates (i.e., sediment toxicity test results or benthic community metrics). Sediment chemistry

data are significantly but weakly correlated with sediment toxicity or benthic community metric variables (Sections 4.2.1). This is corroborated by multivariate statistics (Sections 4.2.2 and 4.2.3), which indicate that sediment chemistry-based factors are related to benthic response variables. MLR models perform poorly, so it is unreasonable to use those models to predict effects on invertebrates.

Habitat variables were often significantly related to benthic response variables, although they tended to be less than or equally important to sediment chemical factors. Similarly, there were no apparent relationships between OC-normalized SEM-AVS (Section 4.3.2) or 34-PAH sum TUs (Section 4.3.3) and sediment toxicity endpoints.

The comparison of LPRSA sediment chemistry data to T20 and T50 values resulted in every SQT location in the LPRSA exceeding the T50 value for at least one chemical. T20 and T50 values cannot reliably predict sediment toxicity in the LPRSA because they are conservative to the point of predicting toxic impacts at 100% of LPRSA SQT locations (Section 4.2.4). There is a clear bias toward type II errors. However, the results of the T20 and T50 screen, although contrary to other sediment chemistry analyses presented in this LOE, will be carried forward into the WOE analysis (Section 6.1 of the BERA main text), consistent with USEPA guidance (USEPA 2015a, b).

5 References

Alden RW, Dauer DM, Ranasinghe JA, Scott LC, Llanso RJ. 2002. Statistical verification of the Chesapeake Bay benthic index of biotic integrity. Environmetrics 13:473-498.

Allen HE, Gongmin F, Boothman W, DiToro DM, Mahony JD. 1991. Draft analytical method for determination of acid volatile sulfide in sediment. Office of Science and Technology, US Environmental Protection Agency, Washington, DC.

Anderson BS, Hunt JW, Phillips BM, Fairey R, Roberts CA, Oakden JM, Puckett HM, Stephenson M, Tjeerdema RS, Long ER, Wilson CJ, Lyons JM. 2001. Sediment quality in Los Angeles Harbor, USA: a triad assessment. Environ Toxicol Chem 20(2):359-370.

Aqua Survey. 2005. Taxonomic identification of benthic invertebrates from sediment collected in the lower 17 miles of the Lower Passaic River in support of the Lower Passaic Restoration Project for NJDOT/OMR. Aqua Survey, Inc., Flemington, NJ.

ASTM. 2007. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. Method E1706-00. Annual book of ASTM standards, vol 11.06. American Society for Testing and Materials, West Conshohocken, PA.

ASTM. 2008. Standard test method for measuring the toxicity of sediment-associated contaminants with estuarine and marine invertebrates. Method E1367-03. Annual book of standards, Vol. 11.05, Biological effects and environmental fate; biotechnology; pesticides. American Society for Testing and Materials, West Conshohocken, PA.

ASTM. 2010. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. Method E1706-00 (reapproved 2010). Annual book of ASTM standards, vol 11.06. American Society for Testing and Materials, West Conshohocken, PA.

Belsley DA, Kuh E, Welsch RE. 1980. Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, Inc., New York, NY.

Bode RW, Novak MA, Abele LE, Heitzman DL, Smith AJ. 2002. Quality assurance work plan for biological stream monitoring in New York State. Bureau of Water Assessment and Management, New York State Department of Environmental Conservation, Albany, NY.

Bollen KA, Jackman RW. 1990. Regression diagnostics: an expository treatment of outliers and influential cases. In: Fox J, Long S, eds, Modern Methods of Data Analysis. Sage Publications, Newbury Park, CA, pp. 257-291.

Breneman D, Richards C, Lozano S. 2000. Environmental influences on benthic community structure in a Great Lakes embayment. J Great Lakes Res 26(3):287-304.

Butler MJ. 1988. In situ observations of bluegill (*Lepomis macrochirus* Raf.) foraging behavior: the effects of habitat complexity, group size, and predators. Copeia 1988(4):939-944.

Canfield TJ, Kemble NE, Brumbaugh WG, Dwyer FJ, Ingersoll CG, Fairchild JF. 1994. Use of benthic invertebrate community structure and the sediment quality triad to evaluate metal-contaminated sediment in the Upper Clark Fork River, Montana. Environ Toxicol Chem 13(12):1999-2012.

Connell JH. 1978. Diversity in tropical rain forests and coral reefs. Science 199(4335):1302-1310.

Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR, Santore RC. 2001. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. Environ Toxicol Chem 20(10):2383-2396.

Dinnel. 2010. Lower Passaic River restoration project: sediment testing quality assurance review. Final report. Dinnel Marine Resources, Anacortes, WA.

DiToro DM, McGrath JA, Hansen DJ, Berry WJ, Paquin PR, Mathew R, Wu KB, Santore RC. 2005. Predicting sediment metal toxicity using a sediment biotic ligand model: methodology and initial application. Environ Toxicol Chem 24(10):2410-2427.

Dziak JJ, Coffman DL, Lanza ST, Li R. 2012. Personal communication Sensitivity and specificity of information criteria. The Methodology Center, Pennsylvania State University, State College, Pennsylvania. June 27, 2012.

EPA. 2003. The biotic ligand model: technical support document for its application to the evaluation of water quality criteria for copper. Health and Ecological Criteria Division, Office of Science and Technology, US Environmental Protection Agency, Washington, DC.

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn CG, Smorong D, Lindskoog R. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ Toxicol Chem 21(9):1993–2005.

Fox J, Weisberg S. 2011. An R Companion to Applied Regression. 2nd ed. Sage Publications, Thousand Oaks, CA.

Gage JD, Hughes DJ, Gonzalez Vecino JL. 2002. Sieve size influence in estimating biomass, abundance and diversity in samples of deep-sea macrobenthos. Mar Ecol Prog Ser 225:97-107.

Germano & Associates. 2005. Final report: sediment profile imaging survey of sediment and benthic habitat characteristics of the Lower Passaic River. Lower Passaic River Restoration Project. Germano & Associates, Inc., Bellevue, WA.

Gorham J, Alevizon W. 1989. Habitat complexity and the abundance of juvenile fishes residing on small scale artificial reefs. Bull Mar Sci 44(2):662-665.

Gray JS, Pearson TH. 1982. Objective selection of sensitive species indicative of pollution-induced change in benthic communities. I. Comparative methodology. Mar Ecol Prog Ser 9:111-119.

Hammerschmidt CR, Burton GA, Jr. 2010. Measurements of acid volatile sulfide and simultaneously extracted metals are irreproducible among laboratories. Environ Toxicol Chem 29(7):1453-1456.

Harrell FE, Jr. 2016. Hmisc: Harrell miscellaneous. R package version 3.17-4. Updated November 1, 2016. Available from: https://cran.r-project.org/web/packages/Hmisc/index.html.

Heck KL, Jr, Wetstone GS. 1977. Habitat complexity and invertebrate species richness and abundance in tropical seagrass meadows. J Biogeog 4(2):135-142.

Hilsenhoff WL. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol 20(1):31-40.

Hilsenhoff WL. 1998. A modification of the biotic index of organic stream pollution to remedy problems and permit its use throughout the year. Great Lakes Entomol 31(1):1-12.

Ingersoll CG, Kunz JL, Hughes JP, Wang N, Ireland DS, Mount DR, Hockett JR, Valenti TW, Jr. 2015. Relative sensitivity of an amphipod *Hyalella azteca,* a midge *Chironomus dilutus,* and a unionid mussel *Lampsilis siliquoidea* to a toxic sediment. Environ Toxicol Chem 34(5):1134-1144.

James G, Witen D, Hastie T, Tibshirani R. 2013. An Introduction to Statistical Learning with Applications in R. Springer, New York, NY.

Jost L. 2006. Entropy and diversity. OIKOS 113(2):363-375.

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.

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.

Long ER, Hong CB, Severn C. 2001. Relationships between acute sediment toxicity in laboratory tests and abundance and diversity of benthic infauna in marine sediments: a review. Environ Toxicol Chem 20(1):46-60.

MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39(5):20-31.

Moffatt & Nichol. 2013. LPR/NB RI/FS modeling program. Status of the CPG's sediment transport model. Moffatt & Nichol Engineers, Seattle, WA.

Molles MC, Jr. 2005. Ecology: concepts and applications. 3rd ed. McGraw-Hill, New York, NY.

Mount DR. 2011. Personal communication (email from D. Mount, USEPA, to D. Ireland, Great Lakes National Program, re: suggested requirements and performance criteria for laboratories conducting sediment toxicity tests). US Environmental Protection Agency, Duluth, MN. November 23, 2011.

Newman MC. 1998. Fundamentals of ecotoxicology. 1 ed. Ann Arbor Press, Ann Arbor, MI.

Nilsson HC, Rosenberg R. 2000. Succession in marine benthic habitats and fauna in response to oxygen deficiency: analysed by sediment profile-imaging and by grab samples. Mar Ecol Prog Ser 197:139-149.

NOAA. 2013. Environmental data retreived from NOAA Query Manager and MARPLOT applications, data, and maps database [online]. National Ocean Service, National Oceanic and Atmospheric Administration, Washington, DC. Available from: http://response.restoration.noaa.gov/environmental-restoration/environmental-assessment-tools/query-manager-marplot-data-maps.html.

Paquin PR, Gorsuch JW, Apte S, Batley GE, Bowles KC, Campbell PGC, Delos CG, Di Toro DM, Dwyer RL, Galvez F, Gensemer RW, Goss GG, Hogstrand C, Janssen CR, McGeer JC, Naddy RB, Playle RC, Santore RC, Schneider U, Stubblefield WA, Wood CM, Wu KB. 2002. The biotic ligand model: a historical overview. Comp Biochem Physiol, Part C 133:3-35.

Pearson TH, Rosenberg R. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanogr Mar Biol Ann Rev 16:229-311.

Peet RK. 1974. The measurement of species diversity. Ann Rev Ecol Syst 5:285-307.

Peet RK. 1975. Relative diversity indices. Ecology 56:496-498.

Phillips BM, Hunt JW, Anderson BS, Puckett HM, Fairey R, Wilson CJ, Tjeerdema R. 2001. Statistical significance of sediment toxicity test results: threshold values derived by the detectable significance approach. Environ Toxicol Chem 20(2):371-373.

Phipps GL, Mattson VR, Ankley GT. 1995. Relative sensitivity of three freshwater benthic macroinvertebrates to ten contaminants. Arch Environ Contam Toxicol 28:281-286.

Pielou EC. 1966. The measurement of diversity in different types of biological collections. J Theoret Biol 13:131-144.

Quinn TP, Peterson NP. 1996. The influence of habitat complexity and fish size on over-winter survival and growth of individually marked juvenile coho salmon *Oncorhynchus kisutch*) in Big Beef Creek, Washington. Can J Fish Aquat Sci 53:1555-64.

R Core Team. 2015. R: A language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

R Core Team. 2016. R: A language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Relyea CD, Minshall GW, Danehy RJ. 2012. Development and validation of an aquatic fine sediment biotic index. Environ Manag 49:242-252.

Rényi A. 1961. On measures of entropy and information. Proc Fourth Berkeley Symp on Math Statist and Prob 1:547-561.

Rosenberg R. 2001. Marine benthic faunal successional stages and related sedimentary activity. Sci Marina 65(Suppl. 2):107-119.
Rosseel Y. 2012. An R package for structural equation modeling. J Stat Softw 48(2):1-36.
Santore RC, Di Toro DM, Paquin PR, Allen HE, Meyer JS. 2001. Biotic ligand model of the acute toxicity of metals. 2. Application to acute copper toxicity in freshweater fish and *Daphnia*. Environ Toxicol Chem 20(10):2397-2402.
Schlekat C, McGee BL, Reinharz E, Velinsky D, Wade TL. 1994. Tidal river sediments in the Washington, DC area. III. Biological effects associated with sediment contamination. Estuaries 17(2):334-344.
Shannon CE. 1948. A mathematical theory of communication. Bell Sys Tech J 27(July and October):379-423 and 623-656.
Simpson EH. 1949. Measurement of diversity. Nature 163:688.
Sousa WP. 1980. The responses of a community to disturbance: the importance of successional age and species' life histories. Oecologia 45:72-81.
Swartz RC, Schultz DW, Ditsworth GR, DeBen WA, Cole FA. 1985. Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. In: Boyle TT, ed, Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems. STP 865. American Society for Testing and Materials, Philadelphia, PA, pp. 152-175.
Swartz RC, Cole FA, Lamberson JO, Ferraro SP, Schults DW, Deben WA, Li HL, Ozretich RJ. 1994. Sediment toxicity, contamination and amphipod abundance at a DDT- and dieldrin-contaminated site in San Francisco Bay. Environ Toxicol Chem 13(6):949-962.
Taniguchi H, Tokeshi M. 2004. Effects of habitat complexity on benthic assemblages in a variable environment. Freshw Biol 49:1164-1178.
USEPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. Second Edition. EPA/600/R-99/064. US Environmental Protection Agency, Washington, DC.
USEPA. 2001a. Methods for collection, storage and manipulation of sediments for chemical and toxicological analyses: technical manual. EPA-823-B-01-002. Office of Water, US Environmental Protection Agency, Washington, DC.
USEPA. 2001b. The role of screening-level risk assessments and refining contaminants of concern in baseline ecological risk assessments. ECO Update bulletin. Publication 9345.0-14, EPA 540/F-01/014. Office of Solid Waste and Emergency Response, US Environmental Protection Agency, Washington, DC.
USEPA. 2002. Role of background in the CERCLA cleanup program. OSWER 9285.6-07P. Office of Solid Waste and Emergency Response, Office of Emergency and Remedial Response, US Environmental Protection Agency, Washington, DC.
USEPA. 2003. Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: PAH mixtures. EPA-600-R-02-013. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

USEPA. 2005a. Contaminated sediment remediation guidance for hazardous waste sites. OSWER 9355.0-85. EPA-540-R-05-012. US Environmental Protection Agency, Office of Solid Waste and Emergency Response Washington, DC.

USEPA. 2005b. Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: metal mixtures (cadmium, copper, lead, nickel, silver and zinc. EPA/600/R-02-011. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

USEPA. 2011. Storet/WQX Data Warehouse [online database]. US Environmental Protection Agency, Washington, DC. Available from: http://www.epa.gov/storet/.

USEPA. 2015a. Sediment quality triad methodology. US Environmental Protection Agency comments on the Draft Lower Passaic River Study baseline ecological risk assessment submitted June 13, 2014. US Environmental Protection Agency, Region 2.

USEPA. 2015b. Technical memorandum, Lower Passaic River Restoration Project guidance for revising the sediment quality triad (SQT) analysis. US Environmental Protection Agency, Region 2.

USEPA. 2016. National coastal assessment data [online]. US Environmental Protection Agency, Washington, DC. Updated February 21, 2016. [Cited July 2016.] Available from: https://archive.epa.gov/emap/archive-emap/web/html/index-78.html.

USEPA. 2017. Personal communication (email from J. LaPoma, USEPA, to R. Law, de maximis, regarding USEPA review of CPG's *2017 R2-CPG Telecon Summary - Action Item Response SQT_1 to EPA*). US Environmental Protection Agency, New York, NY. November 16, 2017.

USEPA, USACE. 1998. Evaluation of dredged material proposed for discharge in waters of the U.S.- testing manual: Inland Testing Manual. EPA-823-B-98-004. US Environmental Protection Agency and US Army Corps of Engineers, Washington, DC.

USEPA REMAP. 1993. Sediment toxicity tests using *Ampelisca abdita* in 1990 and 1993 [online]. USEPA Regional Environmental Monitoring & Assessment Program, National Coastal Assessment-Northeast/New Jersey Coast. [Cited May 14, 2009.] Available from: http://www.epa.gov/emap/nca/html/about.html.

USEPA REMAP. 1998. Sediment toxicity tests using *Ampelisca abdita* in 1998 [online]. USEPA Regional Environmental Monitoring & Assessment Program, National Coastal Assessment-Northeast/New Jersey Coast. [Cited May 14, 2009.] Available from: http://www.epa.gov/emap/nca/html/about.html.

USEPA REMAP. 2002a. Sediment toxicity tests using *Ampelisca abdita* in 2000 and 2002 [online]. USEPA Regional Environmental Monitoring & Assessment Program, National Coastal Assessment-Northeast/New Jersey Coast. [Cited May 14, 2009.] Available from: http://www.epa.gov/emap/nca/html/about.html.

USEPA REMAP. 2002b. Taxonomic identification and biomass of benthic invertebrates from numerous stations along New Jersey coast in 2000 and 2002 [online]. USEPA Regional Environmental Monitoring & Assessment Program, National Coastal Assessment-Northeast/New Jersey Coast. [Cited May 14, 2009.] Available from: http://www.epa.gov/emap/nca/html/about.html.

USEPA REMAP. 2009. Region 2 recent and current REMAP projects [online]. USEPA Regional Environmental Monitoring & Assessment Program. Updated 5/11/09. [Cited 1/28/09.] Available from: http://www.epa.gov/emap/remap/html/two/index.html.

USGS. 2012. USGS 01390000 Passaic River at Garfield NJ: streamflow and temperature. National Water Information System Web Interface [online]. US Geological Survey, West Trenton, NJ. [Cited 8/2/12.] Available from: http://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=01390000.

Weisberg SB, Ranasinghe JA, Dauer DM, Schaffner LC, Diaz RJ, Frithsen JB. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. Estuaries 20:149-158.

Wenning RJ, Batley GE, Ingersoll CG, Moore DW, editors. 2005. Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments. Society of Environmetnal Toxicology and Chemistry (SETAC), Pensacola, FL.

Windward. 2009. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Quality Assurance Project Plan: Surface sediment chemical analyses and benthic invertebrate toxicity and bioaccumulation testing. Final. Prepared for Cooperating Parties Group, Newark, New Jersey. October 8, 2009. Windward Environmental LLC, Seattle, WA.

Windward. 2010. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Spring and summer 2010 benthic invertebrate community surveys. Addendum to the Quality Assurance Project Plan: Surface sediment chemical analyses and benthic invertebrate toxicity and bioaccumulation testing. Addendum no. 1. Final. Prepared for Cooperating Parties Group, Newark, NJ. May 17, 2010. Windward Environmental LLC, Seattle, WA.

Windward. 2012. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Background and reference conditions addendum to the Quality Assurance Project Plan: Surface sediment chemical analyses and benthic invertebrate toxicity and bioaccumulation testing. Addendum No. 5. Final. Prepared for Cooperating Parties Group, Newark, NJ. October 26, 2012. Windward Environmental LLC, Seattle, WA.

Windward. 2014. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Fall 2009 benthic invertebrate community survey and benthic field data collection report for the Lower Passaic River Study Area. Final. Prepared for Cooperating Parties Group, Newark, NJ. Submitted to USEPA January 6, 2014. Windward Environmental LLC, Seattle, WA.

Windward. 2018a. 2012 sediment toxicity reference data for the Lower Passaic River Study Area. Final. Submitted to USEPA October 11, 2018. Prepared for Cooperating Parties Group, Newark, NJ. Windward Environmental LLC, Seattle, WA.

Windward. 2018b. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Fall 2009 sediment toxicity test data for the Lower Passaic River Study Area. Final. Prepared for Cooperating Parties Group, Newark, NJ. Windward Environmental LLC, Seattle, WA.

Windward. Draft. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. 2012 benthic invertebrate community reference data for the Lower Passaic River Study Area. Draft. Prepared for the Cooperating Parties Group, Newark, NJ. Submitted to USEPA October 18, 2018. Windward Environmental LLC, Seattle, WA.

Windward, AECOM. 2015. Lower Passaic River Restoration Project. Data usability and data evaluation plan for the Lower Passaic River Study Area risk assessments. Final. Prepared for Cooperating Parties Group, Newark, NJ. Windward Environmental LLC, Seattle, WA; AECOM, Inc., Westford, MA.

Windward, AECOM. Draft. Lower Passaic River Restoration Project. Revised risk analysis and risk characterization plan for the Lower Passaic River Study Area. Revised draft. Prepared for Cooperating Parties Group, Newark, NJ. Submitted to USEPA October 29, 2013. Windward Environmental LLC, Seattle, WA; AECOM, Inc., Westford, MA.

Wood C. 2012. Biotic ligand model development for Zn and Pb in estuarine and marine waters: interactive influences of salinity and dissolved organic carbon. Prepared for Research Council of Canada, Natural Sciences and Engineering. Department of Biology, McMaster University, Hamilton, Ontario, Canada.

Zar JH. 1996. Biostatistical analysis. 3rd ed. Prentice Hall, Upper Saddle River, NJ.