APPENDIX B. STATISTICAL ANALYSES

Table of Contents

Tables B-ii

Figures B-iii

Acronyms B-iv

B1 Introduction and Methods B-1

B1.1 Choosing the Distributional Form for Calculating 95UCLs B-1

B1.1.1 Goodness-of-fit test B-1

B1.1.2 Probability plots B-2

B1.1.3 Outlier tests B-3

B1.1.4 Boxplots B-3

B1.2 Empirical Cumulative Distribution Plots B-4

B1.3 Variance Components Analysis B-5

B2 Sediment B-7

B2.1 Composite Surface Sediment (0–10-cm) Samples B-7

B2.1.1 Data summary B-7

B2.1.2 95UCL calculations B-7

B2.1.3 Arsenic discussion B-11

B2.2 Composite Intertidal Surface Sediment (0–45-cm) samples B-13

B2.2.1 Potential clamming areas B-13

B2.2.2 Beach play areas B-17

B3 Surface Water B-23

B3.1 Distribution of Passive Sampler Results B-23

B3.2 Evaluation of Sampling Variance B-24

B3.3 Power and Sample Size B-26

B4 Fish and Crab Tissue B-27

B4.1 Influence of Non-Detects B-27

B4.2 95UCL Calculations B-28

B4.3 Statistical Comparisons Between Concentrations of Total PCBs in 2017 Samples (Pre-Design Studies Dataset) and 2007 Samples B-32

B4.3.1 English sole fillet and whole-body tissues B-33

B4.3.2 Shiner surfperch B-34

B4.3.3 Graceful crab B-35

B4.4 Relationship Between PCB Aroclors and Congeners in Fish and Crab Tissues B-36

B4.5 Power and Sample Size B-38

B5 Clam Tissue B-39

B5.1 Influence of Non-Detects B-39

B5.2 95UCL Calculations B-40

Tables

Table B1-1.	Expected means squares for VCA of a single (random) factor and one level of replication in a balanced design	B-6
Table B2-1.	Goodness-of-fit and variance summary statistics for COCs in 0–10-cm sediment composite samples	B-8
Table B2-2.	Summary statistics in potential clamming areas for intertidal (0–45-cm) sediment composites	B-14
Table B2-3.	VCA of potential clamming area composite samples	B-16
Table B2-4.	Summary statistics in beach play areas for intertidal (0–45-cm) sediments	B-19
Table B2-5.	Results of VCA for beach play area composites.	B-20
Table B2-6.	Effect of field duplicates on means and 95UCLs at Beaches 1 and 6	B-22
Table B3-1.	ANOVA table for comparison of total PCBs in passive samplers between two locations and two baseline years (2017 and 2018)	B-25
Table B3-2.	Results of VCA for total PCB passive sampler data	B-25
Table B3-3.	Summary statistics for sum of PCB congeners from PE samplers deployed in the LDW	B-26
Table B4-1.	Goodness-of-fit and relative variance summary for COCs in baseline fish and crab tissues	B-28
Table B4-2.	Summary of total PCB results in English sole fillet and whole-body tissues for baseline and RI datasets	B-33
Table B4-3.	ANOVA table for comparison of total PCBs in English sole tissue between 2007 and 2017	B-34
Table B4-4.	Summary of total PCB results in shiner surfperch whole-body tissues for 2007 and 2017 datasets	B-34
Table B4-5.	ANOVA table for comparison of total PCB Aroclor data in shiner surfperch samples	B-35
Table B4-6.	Summary of total PCB Aroclor results in graceful crab tissues for the 2007 and 2017 datasets	B-36
Table B4-7.	ANOVA table for comparison of total PCB Aroclor data in graceful crab samples (Reach 1 only)	B-36
Table B4-8.	Number and type of fish and crab tissue samples with PCB Aroclors and Congeners	B-37
Table B4-9.	Expected MDDs for comparisons between baseline and future monitoring for COCs and species/tissue types with TTLs	B-39
Table B5-1.	Goodness-of-fit and variance statistics for risk drivers in clam tissue composite samples	B-43

Figures

Figure B1-1. Example probability plots for a skewed dataset that does not follow a normal distribution (left) but does follow a lognormal distribution (right) B-2

Figure B1-2. Example boxplot with labels of the distributional characteristics represented by the different parts of the boxplot B-4

Figure B1-3. Example ECDF plot B-5

Figure B2-1a. Probability plots of total PCB (sum of Aroclors) and cPAH TEQ results in composite samples from 0–10-cm sediments B-9

Figure B2-1b. Probability plots of dioxin/furan TEQ and arsenic results in composite samples from 0–10-cm sediments B-10

Figure B2-2. Empirical cumulative distribution curves for arsenic concentrations in OSV *Bold* natural background and LDW baseline composite sediment datasets, and theoretical curve for best-fit distribution to the baseline data B-12

Figure B2-3. Results for the LDW-wide clamming area intertidal (0–45-cm) sediments B-15

Figure B3-1. Normal probability plot of station residuals for the baseline passive sampler dataset (n=35) B-24

Figure B3-2. Relationship between replication within each station/depth for future sampling event and scaled MDD B-27

Figure B4-1a. Normal probability plots of baseline total PCB Aroclors (µg/kg ww) in fish and crab composite tissue samples B-29

Figure B4-1b. Normal probability plots of baseline total PCB Aroclors (µg/kg ww) in fish and crab composite tissue samples B-30

Figure B4-2a. Normal probability plots of baseline dioxin/furan TEQ (ng/kg ww) in fish and crab composite tissue samples B-31

Figure B4-2b. Normal probability plots of baseline dioxin/furan TEQ (ng/kg ww) in fish and crab composite tissue samples B-32

Figure B4-3. Plot of Aroclors vs. congeners for baseline fish and crab tissues B-38

Figure B5-1. Probability plots of inorganic arsenic (mg/kg ww) results in clam tissue composite samples B-40

Figure B5-2. Probability plots of cPAH TEQ (µg/kg ww) results in clam tissue composite samples B-41

Figure B5-3. Probability plots for dioxin/furan TEQs (ng/kg ww) and total PCBs (µg/kg ww) in clam tissue composite samples B-42

Acronyms

95URL	95% upper confidence limit (on the mean)
ANOVA	analysis of variance
COC	contaminant of concern
cPAH	carcinogenic polycyclic aromatic hydrocarbon
CV	coefficient of variation
DQO	data quality objective
DF	detection frequency
dw	dry weight
ECDF	empirical cumulative distribution function
Ecology	Washington State Department of Ecology
ID	identification
IQR	interquartile range
LDW	Lower Duwamish Waterway
MDD	minimum detectable difference
MDL	method detection limit
OLS	ordinary least squares
OSV	ocean survey vessel
PCB	polychlorinated biphenyl
PE	polyethylene
PPCC	probability plot correlation coefficient
QAPP	quality assurance project plan
RAO	remedial action objective
RI	remedial investigation
RL	reporting limit
SCO	sediment cleanup objective
SD	standard deviation
SE	standard error (of the mean)
TEQ	toxic equivalent

TTL	target tissue level
VCA	variance components analysis

B1 Introduction and Methods

This appendix to the *Lower Duwamish Waterway Data Evaluation Report* presents the analytical methods and detailed results of the statistical evaluations that were used to interpret the baseline datasets in light of the data quality objectives (DQOs).

The remainder of this appendix is organized into sections that parallel the structure of the main report, which includes the following sections:

- Section 2 – Sediment
- Section 3 – Surface Water
- Section 4 – Fish and Crab Tissue
- Section 5 – Clam Tissue
- Section 6 – References

The statistical methods that were applied to one or more datasets in later sections are described in Section 1.

B1.1 Choosing the Distributional Form for Calculating 95UCLs

The 95% upper confidence limit (on the mean) (95UCL) is a summary statistic required for many of the Lower Duwamish Waterway (LDW) baseline datasets. The 95UCL for each dataset was calculated using the appropriate parametric equations following identification of the most appropriate distributional form (i.e., normal, log-normal, or gamma). If one of the parametric distributions was not appropriate for a dataset, then a non-parametric approach was required. This process also allowed for identification of any possible outliers in a dataset so that these elevated values could be discussed further.

Each dataset was evaluated using tools in ProUCL 5.1 (EPA 2016) and select packages (e.g., EnvStats (Millard 2013) and ggplot2 (Wickham 2009)) in R (R Core Team 2018)). The statistical tools used during this assessment included probability plots, distributional goodness-of-fit tests, and graphical and formal outlier tests.

B1.1.1 Goodness-of-fit test

A formal goodness-of-fit test was conducted for each chemical dataset individually, and each test was confirmed by patterns observed in the probability plots (discussed below). The best-fitting distribution was identified as the one that passed the goodness-of-fit test and had the highest probability plot correlation coefficient (PPCC). If no distributions provided a reasonable fit to the data, then non-parametric estimates for the 95UCL were required.

For this evaluation, goodness-of-fit testing relied on the significance of the probability plot correlation coefficient (using *EnvStats::gofTest(x, test="ppcc,"*

estimate.params=TRUE) in R) for normal, lognormal, and gamma distributions, with the hypothesized distribution rejected when $p < 0.05$. Once the best distributional fit for a dataset was identified, the 95UCL was calculated in ProUCL 5.1 (EPA 2016).

B1.1.2 Probability plots

Probability plots show the observed quantiles for the dataset on the y-axis vs. the expected quantiles under the theorized distribution on the x-axis (hence the synonymous name "QQ Plot," which stands for quantile-quantile plot). If the theoretical distribution is a reasonable description for the dataset, then this plot should follow an approximately straight line. The best-fit regression line is added to the plots to facilitate interpretation of the goodness-of-fit indicated by these plots. These plots are generated in R using the function *EnvStats::qqPlot(x, estimate.params=TRUE)*. The presence of potential outliers and systematic deviations from the theorized distribution can also be observed on these plots; if present, such outliers and deviations may lead to a formal outlier test, as described in the next section. Figure B1-1 shows example probability plots for a skewed dataset that is poorly described by a normal distribution (i.e., the observed quantiles do not fit a straight line when plotted against the normal quantiles) but adequately described by a lognormal distribution (i.e., the QQ plot follows an approximately straight line).

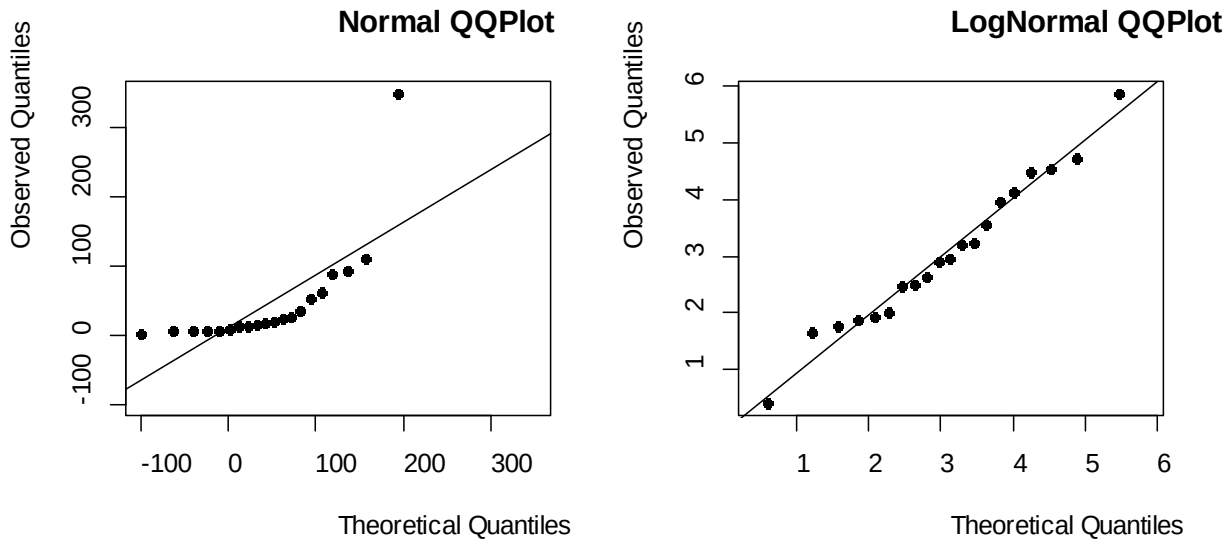


Figure B1-1. Example probability plots for a skewed dataset that does not follow a normal distribution (left) but does follow a lognormal distribution (right)

B1.1.3 Outlier tests

The presence of potential outliers was identified initially through visual inspection of the probability plots. When data points appeared to be extreme (either at the high or low end), a formal outlier test was used. Outlier tests require a parametric assumption for the underlying data; there is no such thing as an outlier for a non-parametric distribution. The two outlier tests used are based on an underlying normal distribution. It is usually the case that the skewness introduced by extreme values can be adequately described by a log-normal or gamma distribution. Alternatively, once extreme values have been removed, the data may be adequately described by a normal distribution, which is the basis for the two outlier tests: Dixon's ($n < 25$, single outliers only) and Rosner's ($n \geq 25$, multiple outliers). Both tests were applied using tools in ProUCL 5.1 (EPA 2016).

B1.1.4 Boxplots

Boxplots (a.k.a. box-and-whisker plots) are used to illustrate the distribution of the data, providing information about the location and spread of the data as well as skewness. Boxplots are especially useful when several are placed side by side. Each boxplot has a shaded/colored rectangle that shows the spread of values between the 1st and 3rd quartiles (i.e., the 25th and 75th percentiles). The height of this rectangle is the interquartile range (IQR), which is simply the value of the third quartile minus the value of the first quartile. The line inside the box indicates the median; the outer brackets (the "whiskers") represent the minimum and maximum values or 1.5 times the IQR from the median, whichever is less; values outside the whiskers are possible extreme values and are shown as individual data points. The median plus and minus 1.5 times the IQR is expected to contain about 98% of a Standard Normal (Gaussian) distribution. Boxplots were generated in R using the function *ggplot + geom_boxplot*. Figure B1-2 is an example boxplot with labels of the distributional characteristics represented by the different parts of the boxplot.

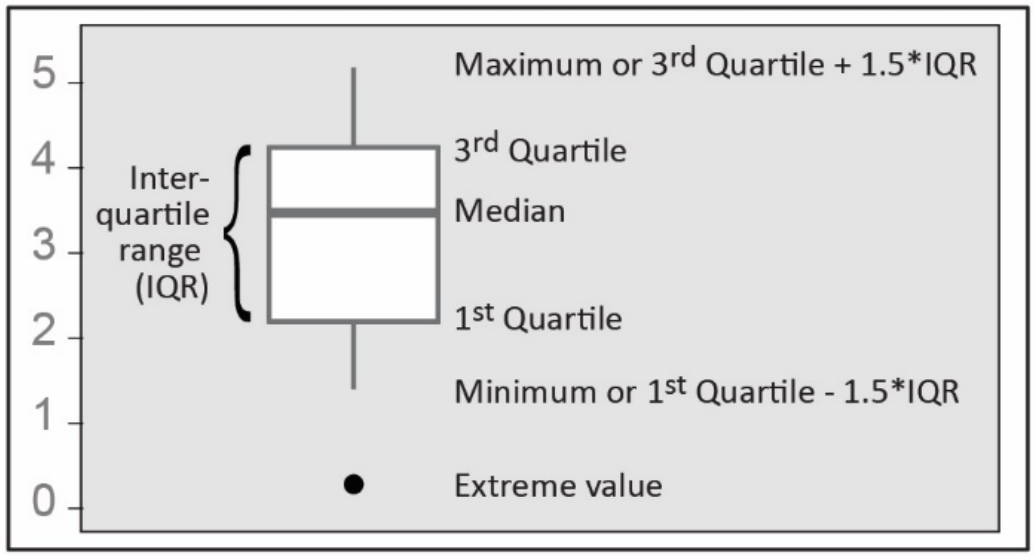
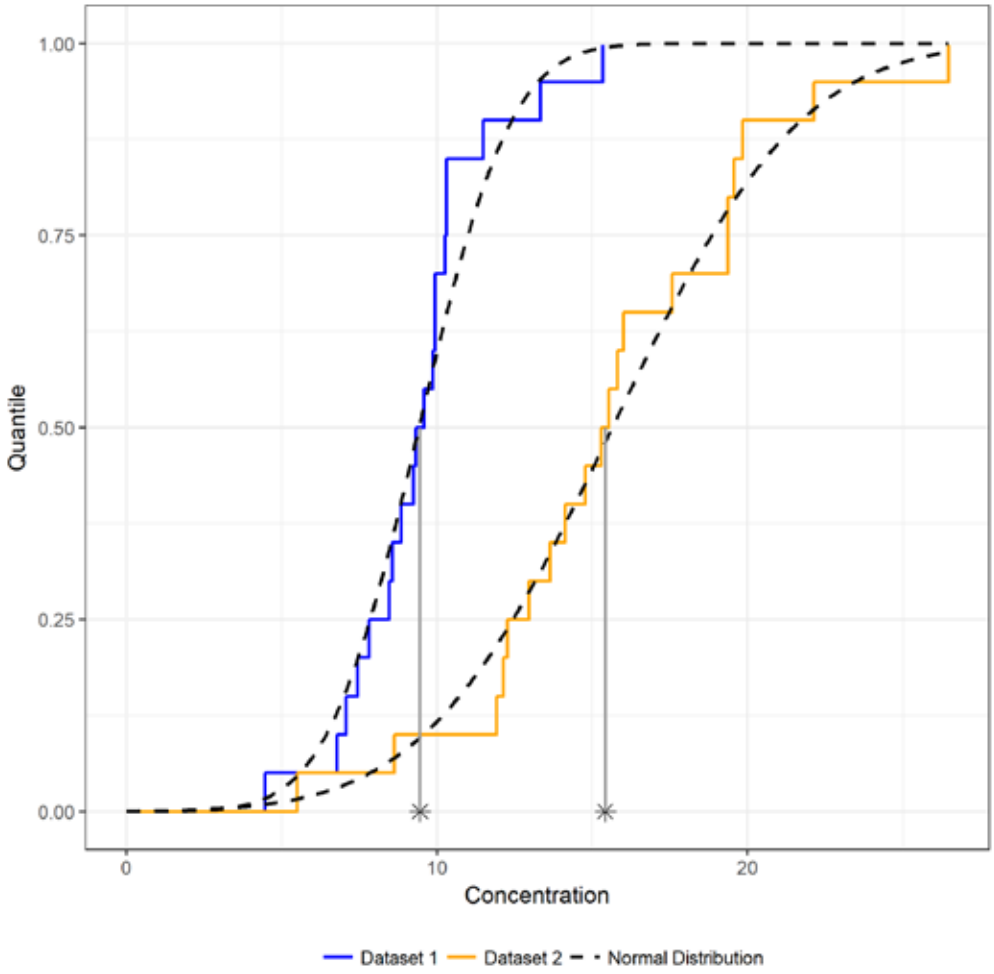


Figure B1-2. Example boxplot with labels of the distributional characteristics represented by the different parts of the boxplot

B1.2 EMPIRICAL CUMULATIVE DISTRIBUTION PLOTS

Useful in illustrating the distribution and skewness in a dataset, empirical cumulative distribution function (ECDF) plots display the percentiles or cumulative probabilities for each observation in a dataset (Figure B1-3). Each distribution is shown as a step function, with a step up at each unique concentration. These plots provide visualization of where an individual threshold concentration (e.g., a cleanup level) may fall along the distribution of concentrations in the LDW dataset, and can also be used to compare multiple datasets (e.g., in cases where the remedial action objective [RAO] cleanup levels are background based, the ECDF for the LDW baseline dataset can be contrasted with the ECDF for the ocean survey vessel (OSV) *Bold* background dataset (Figure B2-1) (DMMP 2009). ECDF plots readily allow the interpretation of certain characteristics of data distributions. When two curves are shown on the same plot, the curve further to the right has higher concentrations (e.g., Dataset 2 in Figure B1-3); steeper curves have less variance (e.g., Dataset 1 in Figure B1-3) and specific percentiles can be readily identified (e.g., the median concentration is the concentration on the x-axis that coincides with a cumulative probability of 0.5 on the y-axis for a particular curve, as indicated by asterisks on the x-axis in Figure B1-3). ECDF plots were generated in R using the function *ggplot + stat_ecdf.*



Note: Asterisks indicate the median concentration for the two datasets. The normal distribution curves use mean and variance estimated from each dataset.

Figure B1-3. Example ECDF plot

B1.3 VARIANCE COMPONENTS ANALYSIS

An analysis of the variance components was used to investigate the relative importance of different sources of variability of the total sampling variance (i.e., analytical, compositing, or spatial variability). The datasets for which variance components were analyzed were not explicitly designed with this analysis in mind, so the results reflect only possible patterns in the data collected.

Variance components analysis (VCA) uses the analysis of variance (ANOVA) model to partition the sums of squares into their component parts. In the same way that the sources of variance in an ANOVA model are isolated in hypothesis testing to express the "statistical significance" of a particular feature of the study design, the sources of variance can be expressed as a percent of the total to express the relative importance of the variance of that feature. The VCAs were conducted using *anovaVCA* in R (Schuetzenmeister and Dufey 2018). A conservative estimate of total variance was

used by setting negative variance components estimates to zero *(NegVC = FALSE)*. Negative variance components can arise from the additive model used to partition the sums of squares from the expected variance components, described below.

Table B1-1 shows the theoretical (modeled) expectation for each variance component in a design with multiple levels of one factor that represents a source of variance (e.g., location, or year) and replication within each level of that factor (e.g., field replicates of the beach play composites, or polyethylene (PE) sampler replication at each station within each year). The expected variance components are derived from the observed mean squares by subtraction.

Table B1-1. Expected means squares for VCA of a single (random) factor and one level of replication in a balanced design

Source	Degrees of Freedom	Expected Mean Squares	Observed Mean Squares	Estimated Variance Component
Factor A (e.g., location)	*a* – 1	$\sigma_\varepsilon^2 + n\sigma_A^2$	SSA/*(a-1)*	$\widehat{\sigma_A^2} = \frac{1}{n}\left[\frac{SSA}{a-1} - \frac{SSE}{a(n-1)}\right]$
Within Factor A (e.g., field replication)	*a*(*n*-1)	σ_ε^2	*SSE/a(n-1)*	$\widehat{\sigma_\varepsilon^2} = \left[\frac{SSE}{a(n-1)}\right]$

SSA – sum of squared residuals for Factor A =

$$SSA = \sum_{i=1}^{a}\sum_{j=1}^{n}(\overline{y_{i.}} - \overline{y_{..}})^2$$

SSE – sum of squared residual errors =

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y_{i.}})^2$$

Where *i* indicates the level of Factor A (running from 1 to a), and *j* indicates the individual observation within each level of Factor A (running from 1 to n).

VCA – variance components analysis

For example, in a design that has only one factor with replication, the expected mean squares among the lowest level of replication (e.g., field replicates for the beach play composites) are the estimates of error variance (σ_ε^2). The expected mean squares among the primary factor levels of Factor A (e.g., the independent beach play composites) are the sum of the independent sources of: a) error variance (σ_ε^2) and b) variance among levels of the primary factor ($n\sigma_A^2$) (where *n* is the number of replicates within each level of the factor, and σ_A^2 is the variance among those levels). The mean squared error among levels of Factor A provides an estimate of the total from that variance source (i.e., $\sigma_\varepsilon^2 + n\sigma_S^2$). Hence, the variance due to sampling location only (σ_S^2) is estimated by subtracting σ_ε^2 and dividing by n. This can lead to negative estimates for a variance component (a mathematical possibility but an odd situation, nonetheless) if there is high variability in σ_ε^2, and if the values from the higher level of replication (sampling locations) all overlap with one another.

B2 Sediment

This section provides details of the statistical analyses summarized in Section 2 of the main report for the sediment data collected in February / March and June 2018 per the surface sediment quality assurance project plan (QAPP) (Windward 2018a). The data were presented in the sediment data report (Windward 2018b).

B2.1 COMPOSITE SURFACE SEDIMENT (0–10-CM) SAMPLES

B2.1.1 Data summary

The surface sediment composite sample dataset consisted of 24 samples; each composite sample was composed of 7 grab samples. Composite samples were analyzed for the four risk drivers (total polychlorinated biphenyl [PCB] Aroclors, carcinogenic polycyclic aromatic hydrocarbons [cPAHs], dioxins / furans, and arsenic).

All four risk drivers had 100% detection frequency (DF).[1] Total PCBs were calculated as the sum of detected Aroclors; at least one Aroclor was detected in each sample. Every cPAH compound in each composite sample was detected. For dioxins / furans, non-detected congeners, when included at ½ detection limit, represented ≤ 1% of the total TEQ for most of the samples; in the remaining 6 samples, non-detected congener TEQ contributions ranged from 2 to 34%. The minor presence of non-detects in this dataset did not negatively affect the utility of these data to estimate site-wide mean and 95UCL estimates.

B2.1.2 95UCL calculations

Sediment DQO 1 required that the 95UCL for the site-wide mean be established from this dataset for the four risk drivers. Following the methods described in Section 1.1, the best distributional form for each contaminant of concern (COC) was identified and the 95UCL was calculated (Table 2-2 in main report). Goodness-of-fit and variance summary statistics for the four risk drivers in the composite sediment dataset are shown in Table B2-1 and illustrated in the probability plots (Figures B2-1a, B2-1b).

[1] Total PCBs (sum of Aroclors) was a sum of detected values only. If no Aroclors were detected, then the sum was reported as the highest reporting limit (RL) and U-qualified (not detected at given concentration). For weighted sums (i.e., toxic equivalents [TEQs]), non-detects were included at one-half the RL. If none of the components were detected, the sum of the weighted one-half RLs was reported (and TEQ was U-qualified).

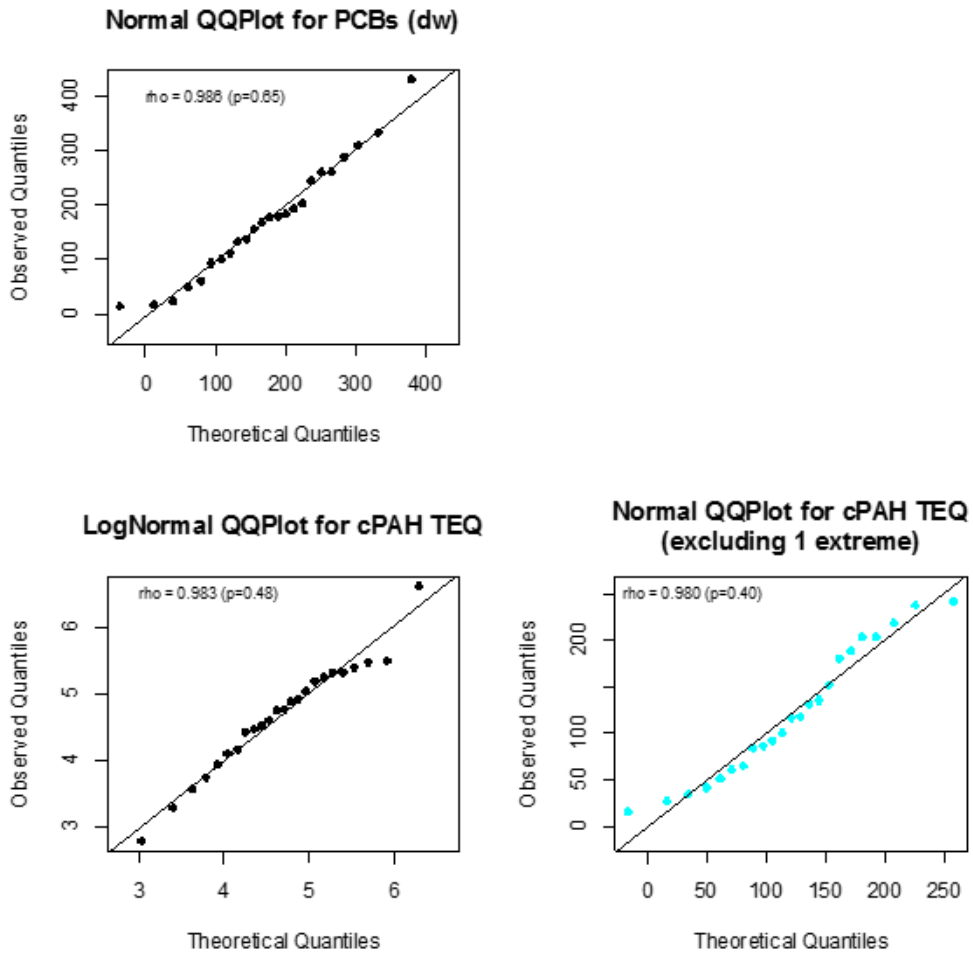
Table B2-1. Goodness-of-fit and variance summary statistics for COCs in 0–10-cm sediment composite samples

COC (units)	Best-fitting Distribution	PPCC[a] (p-value)	CV	Comment
PCB sum of Aroclors (ug/kg dw)	normal	0.986 (p = 0.65)	0.62	The normal distribution is a good fit.
cPAH TEQ (µg/kg dw)	lognormal	0.983 (p = 0.48)	0.98	One elevated influential value present, which skews the dataset.
cPAH TEQ (µg/kg) – exclude outlier	normal	0.98 (p = 0.40)	0.58	Distribution excludes highest value (COMP-02, with concentration of 742 µg/kg). The normal distribution is a good fit.
Dioxin/furan TEQ (ng/kg dw)	gamma	0.986 (p = 0.62)	0.79	Two elevated influential values present, which skew the dataset.
Dioxin/furan TEQ (ng/kg dw) - exclude outliers	normal	0.970 (p=0.18)	0.62	Distribution excludes highest values (COMP-6 and COMP-11, with concentrations of 22.5 and 27.7 ng/kg, respectively). The normal distribution is a good fit.
Arsenic (mg/kg, dw)	lognormal	0.978 (p=0.32)	0.37	One elevated influential value present, which skews the dataset.
Arsenic (mg/kg dw) - exclude outlier	normal	0.994 (p=0.98)	0.26	Distribution excludes highest value (COMP-20, with concentration of 27.2 mg/kg). The normal distribution is a good fit.

[a] PPCC for the best fit distribution for this dataset.

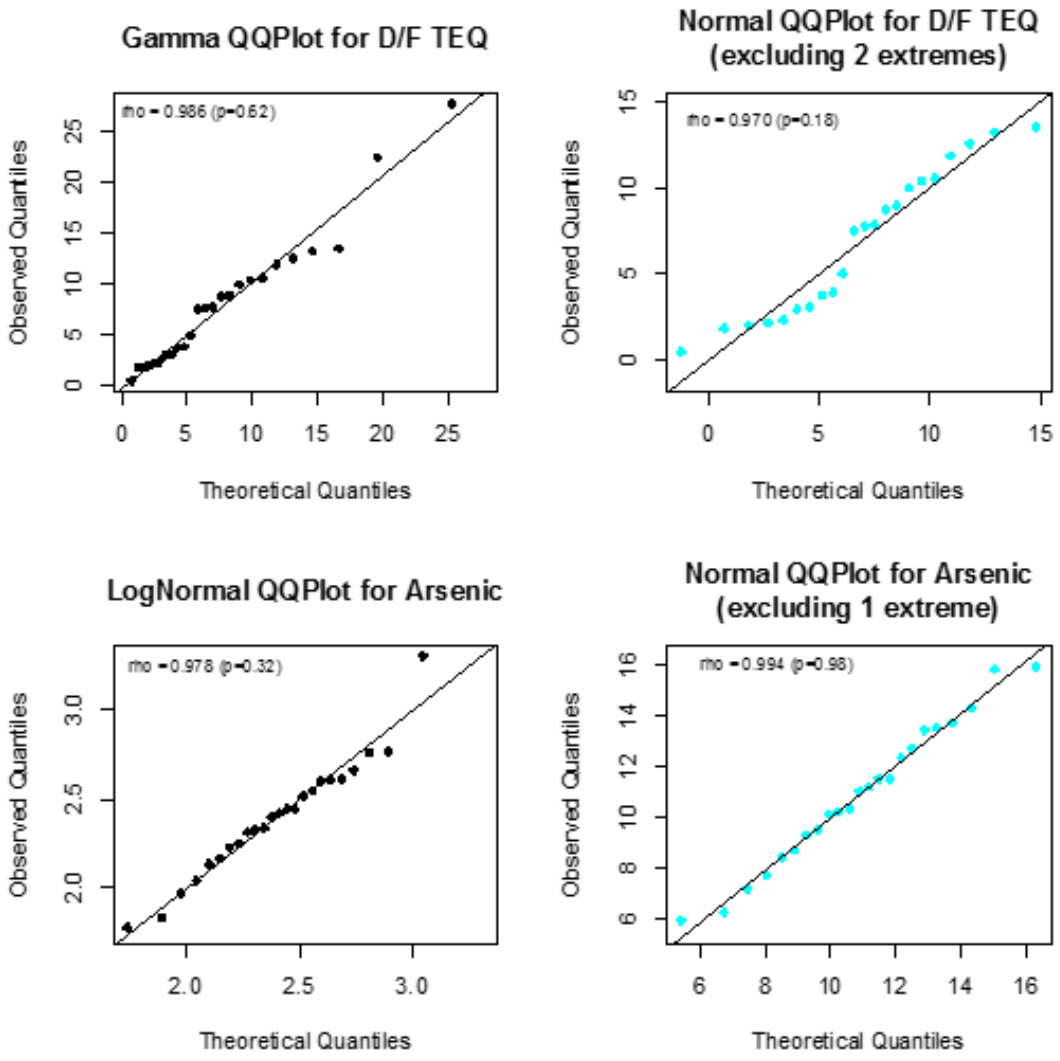
COC – contaminant of concern
CV – coefficient of variation
dw – dry weight
PCB – polychlorinated biphenyl
PPCC – probability plot correlation coefficient
TEQ – toxic equivalent

The data distributions for the cPAH TEQ, dioxin/furan TEQ, and arsenic had one or two elevated influential values each that skewed the data distributions. In order to evaluate the impact of these high values, the best-fit distributions were also evaluated excluding the influential values. When the elevated influential values were excluded (or not present, as in the PCBs dataset), the data distributions were all normally distributed, with coefficients of variation (CVs) of 0.62 or less. This result indicates that, with the exception of one or two individual samples, each of the baseline LDW datasets was well-behaved and had variability that was similar to or better than the assumed variability used to develop the sampling design.



Plots with black dots show the best-fitting distribution for all 24 composite samples. Plots with teal dots show the best-fitting distribution excluding the extreme and influential values that skewed the complete distribution.

Figure B2-1a. Probability plots of total PCB (sum of Aroclors) and cPAH TEQ results in composite samples from 0–10-cm sediments



Plots with black dots show the best-fitting distribution for all 24 composite samples. Plots with teal dots show the best-fitting distribution excluding the extreme and influential values that skewed the distribution of all data.

Figure B2-1b. Probability plots of dioxin/furan TEQ and arsenic results in composite samples from 0–10-cm sediments

With the exception of one or two individual samples, the baseline surface sediment data distributions were all normally distributed, with CVs of approximately 0.6 or less. This is expected to be a reasonable representation of the sampling variability in site-wide sediments after remediation. For normally distributed datasets, the normal *t*-interval is used to calculate the 95UCL for the site-wide mean. The resulting RME expressed as percent of the mean is calculated using Equation 1:

$$\%RME = CV \times \frac{t_{(0.05,n-1)}}{\sqrt{n}} \times 100 \qquad \textbf{Equation 1}$$

When the CV = 0.6, n = 24 composites, and $t_{(0.05,n-1)}$ = 1.714, the %RME is estimated to be 21% of the mean.

B2.1.3 Arsenic discussion

The arsenic results for the composite sediment dataset were evaluated using an expanded statistical approach because, instead of the arsenic cleanup level being risk-based, it was based on the natural background distribution. The ROD established the arsenic RAO 2 cleanup level as the 95UCL of the dataset collected by EPA's OSV *Bold*, which consisted of 70 individual grab samples collected from Puget Sound natural background areas (EPA 2014). Per the ROD (Table 19, footnote e),[2] determination of compliance may be established by one of the following approaches:

- Approach 1 – A direct comparison of the 95UCL of the LDW dataset mean with the 95UCL of the OSV *Bold* dataset mean
- Approach 2 – A statistical comparison of the distribution of the LDW dataset to the OSV *Bold* background dataset

These determinations of compliance can be interpreted as either intending that the post-remedy site should have mean concentrations similar to natural background (Approach 1),[3] or that the entire distribution should be similar to natural background (Approach 2).

There are major differences in the two datasets (i.e., composites from the LDW dataset vs. individual grab samples from Puget Sound in the OSV *Bold* dataset), which influence how compliance, or progress toward compliance, may be appropriately evaluated. For example, the baseline 95UCL may be used to establish whether the mean of the baseline distribution can be expected to be below some bright-line threshold with 95% confidence (similar to Approach 1), but it would be inappropriate to expect entire distributions to be similar because they are of different types of samples (i.e., individual grabs vs. composites). The empirical cumulative distributions of the two datasets (i.e., LDW baseline and OSV *Bold*) are shown in Figure B2-2.

[2] ROD Table 19 is titled *Cleanup levels for PCBs, arsenic, cPAHs, and dioxins/furans in sediment for human health and ecological COCs (RAOs 1, 2, and 4).*

[3] Comparing the 95UCL of one distribution to the 95UCL of another distribution does not allow any probability statements to be made about the relationship between the two means. Instead, when this compliance test is met, there will be at least 95% confidence that the post-remedy site mean is less than the bright-line threshold established by the 95UCL of the OSV *Bold* dataset.

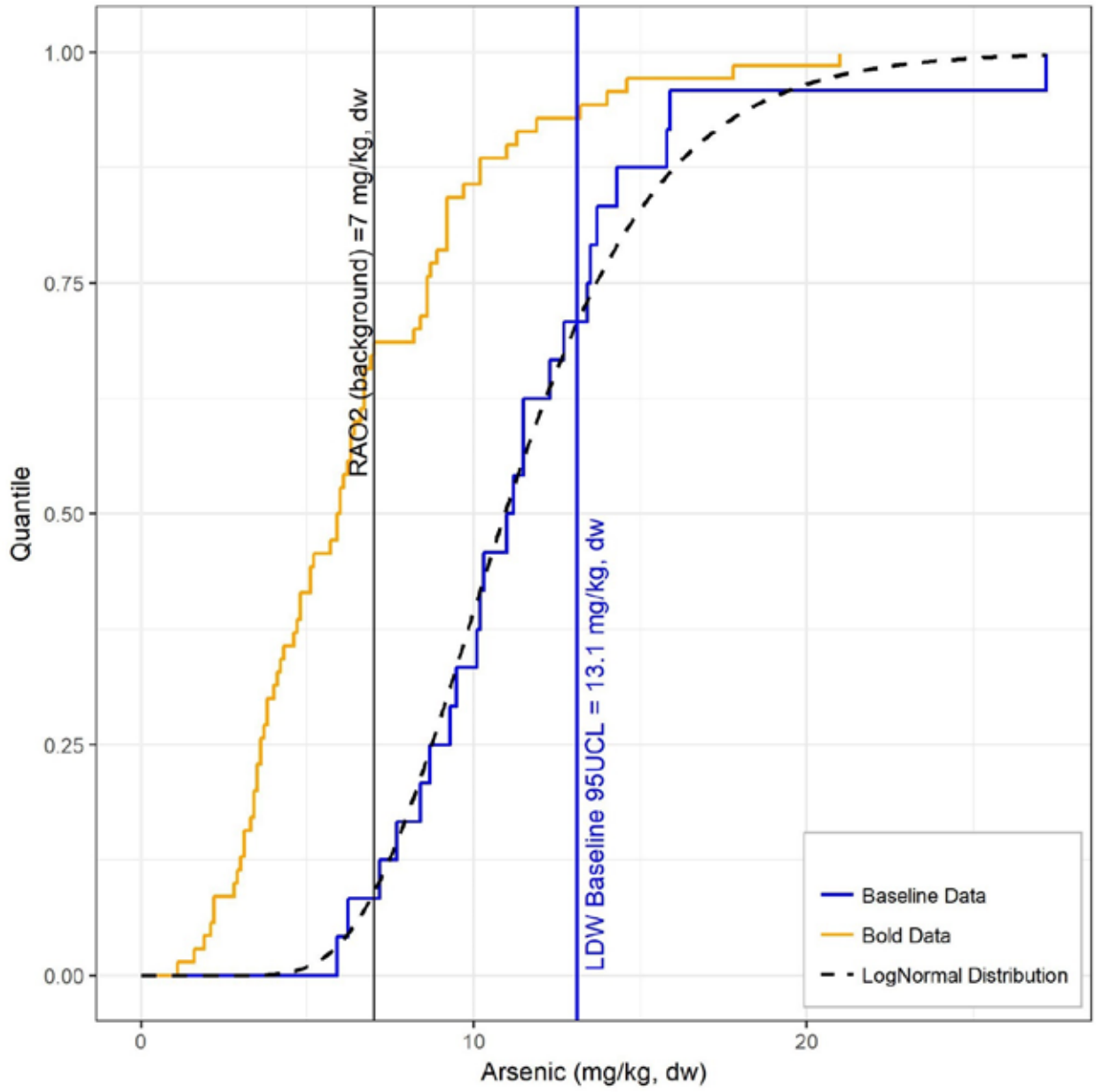


Figure B2-2. Empirical cumulative distribution curves for arsenic concentrations in OSV *Bold* natural background and LDW baseline composite sediment datasets, and theoretical curve for best-fit distribution to the baseline data

The two approaches identified in the ROD (EPA 2014) intending statistical similarity between the LDW and natural background concentrations are more restrictive than Washington State cleanup standards [SCUM II] (Ecology 2015). State standards use the OSV *Bold* natural background dataset differently than does the ROD to establish the sediment cleanup objective (SCO). The Washington State Department of Ecology (Ecology) sets the SCO at a value from the upper tail of the background distribution

(the 90/90 upper tolerance limit [UTL][4]), whereas the ROD sets the cleanup level at a value near the central tendency of the background distribution (the 95UCL) (EPA 2014). The subsequent test for compliance within the state standards is to compare the site mean to the 90/90 UTL as a bright-line threshold. This test can be interpreted as requiring concentrations in the post-remedy site population to be, on average, not more than concentrations in 90% of the natural background population, with 50% confidence.[5] The baseline arsenic mean is 12 mg/kg, which is in between the background 90/90 UTL of 13 mg/kg based on the OSV *Bold* dataset and the value of 11 mg/kg for the OSV *Bold* Plus dataset (Table 10-1, Ecology 2015).

B2.2 COMPOSITE INTERTIDAL SURFACE SEDIMENT (0–45-CM) SAMPLES

B2.2.1 Potential clamming areas

The dataset for sediments from potential clamming areas consisted of 3 site-wide composite samples with 68 samples in each (for a total of 204 grab samples). Composites were analyzed for PCBs, arsenic, cPAHs, and dioxins/furans.

All four risk drivers had 100% DF in this dataset. The constituent compounds for dioxin/furan TEQ and cPAH TEQ were also 100% detected. Total PCBs were calculated as the sum of detected Aroclors; at least one Aroclor was detected in each sample.

B2.2.1.1 95UCL calculations

Sediment DQO 7 (applies to potential clamming area sediment) requires that the 95UCL for the mean of the LDW-wide potential clamming areas be established from this dataset for the four risk drivers. The 95UCL was derived using a *t*-interval for a normally distributed population[6] using Equation 2.

$$95UCL = \bar{X} + t_{(0.05,df=2)} \times SE \qquad \textbf{Equation 2}$$

Where the additional terms are defined as:

SE = standard error, or standard deviation of the mean, calculated from the 3 site-wide composites

df = degrees of freedom, equal to sample size (n) - 1

The summary statistics for the four risk drivers in this dataset are presented in Table B2-2 and Figure B2-3. For a small sample size (e.g., n = 3), it is not unexpected

[4] The Washington State SCO is the 90/90 UTL of natural background, the 90/90 UTL being the concentration at which there exists 90% confidence that 90% of the natural background population will not exceed the limit.

[5] The mean without a UCL is essentially a 50% confidence limit for the mean.

[6] Because each analytical sample represented the potential clamming area-wide mean based on a large number of grab samples (n = 68) per composite, the Central Limit Theorem (CLT) was invoked and normality assumed.

that a random sample would appear to be asymmetrical (e.g., total PCBs and dioxins/furans TEQ in the clamming area sediments, Figure B2-3). This apparent skewness does not automatically refute the normality of a small sample size, and the theoretical underpinning of the CLT is relied upon. The CLT states that the mean (i.e., the physical averaging through compositing) of 68 individual samples should be approximately normally distributed. If the underlying distribution is a skewed distribution rather than a normal distribution (e.g., gamma or lognormal), then the 95UCL provided by Equation 2 will have coverage of the true mean that is less than 95% (i.e., the 95UCL will be too low).

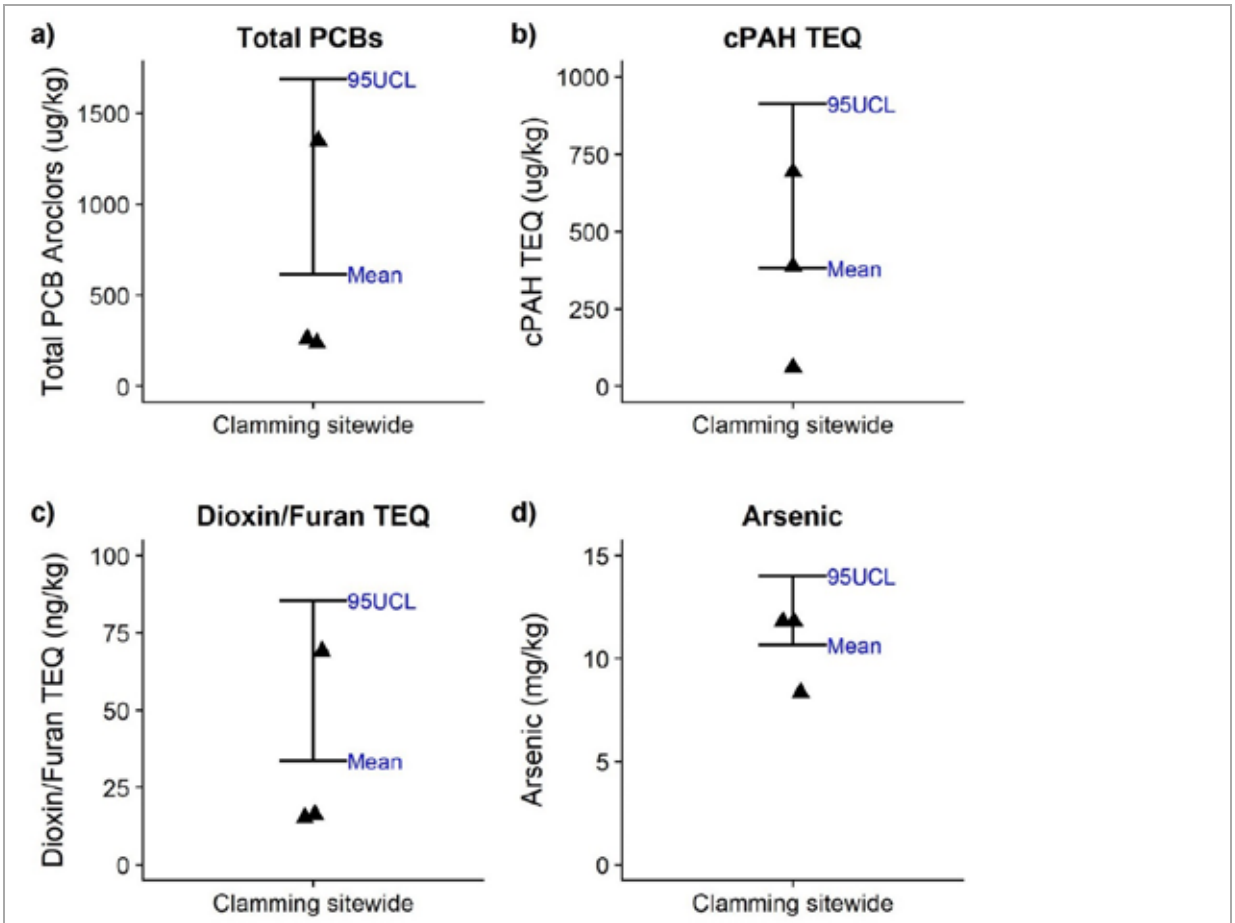
Table B2-2. Summary statistics in potential clamming areas for intertidal (0–45-cm) sediment composites

COC (units)	Potential Clamming Areas Site-wide			
	Mean	CV[a]	SE	95UCL[b]
Total PCB Aroclors (µg/kg)	617	103%	367	1690
cPAH TEQ (µg/kg)	381	83%	182	913
Dioxins/furans TEQ (ng/kg)	33.6	92%	17.8	85.5
Arsenic (mg/kg)	10.7	19%	1.15	14.0

[a] CV = SD/mean

[b] 95UCL calculated using the t-interval (degrees of freedom = 2) for the clamming area composites. These estimates do not use the homogenization duplicates taken for clamming area composite sample 1 (LDW18-IT45-CL-Comp1).

95UCL – 95% upper confidence limit (on the mean)
COC – contaminant of concern
cPAH – carcinogenic polycyclic aromatic hydrocarbon
CV – coefficient of variation
SD – standard deviation
SE – standard error (of the mean)
TEQ – toxic equivalent



Note: For each COC, the three individual sample results are shown (black triangles), along with the mean and 95UCL.

Figure B2-3. Results for the LDW-wide clamming area intertidal (0–45-cm) sediments

B2.2.1.2 Evaluation of sampling variance

The sampling variance is the total variance observed in the dataset; it includes spatial, homogenization, and analytical variance. The sampling variability of results for the clamming area-wide composites was high (CVs greater than 83% for all but arsenic, Table B2-2). Spatial variability is present on multiple spatial scales: The study was designed to capture large-scale spatial variability through the use of the site-wide composites, while micro-scale spatial variability was included in the homogenization triplicates. Large-scale spatial variance can only be estimated by subtracting the within-composite variance through a VCA. The within-composite variance included micro-scale spatial variability, homogenization, and analytical variability; it was captured by the triplicate samples generated from one composite. Three samples were created by subsampling the trays for composite 1 (LDW18-IT45-CL-Comp1) three

times for a triplicate analysis of cPAHs and PCBs.[7] Using VCA (Section 1.3), a relative comparison of the variance among these triplicates to the variance among all clamming area composite samples provided an estimate of the combined micro-scale spatial, homogenization and analytical variances (Table B2-3). For PCB Aroclors, the variability among homogenization triplicates was < 1% of the total variance, indicating very good consistency within the composite tray for this analyte. For cPAHs, variability among homogenization triplicates was 31% of the total variance. For TOC, variability among homogenization triplicates was 49% of the total variance.

Table B2-3. VCA of potential clamming area composite samples

Variance Source	Degrees of Freedom	Sum of Squared Error	Mean Squared Error	Variance Component	% Total for Variance Components
Total PCBs (CV = 128%)[a]					
Total Observed	2.01	na	na	351548	100
Among Composites	2	983019	491509	349904	99.5
Within Composite[b]	2	3289	1644	1644	0.5
cPAH TEQ (CV = 116%)					
Total Observed	2.4	na	na	94361	100
Among Composites	2	240930	120465	65261	69
Within Composite[b]	2	58200	29100	29100	31
Total Organic Carbon (CV = 23%)					
Total Observed	2.6	na	na	0.114	100
Among Composites	2	0.277	0.138	0.058	51
Within Composite[b]	2	0.113	0.056	0.056	49

[a] CV among 6 samples.

[b] The homogenization variance estimated using three subsamples from the composite tray for clamming area sediment composite LDW18-IT45-CL-Comp1.

cPAH – carcinogenic polycyclic aromatic hydrocarbon
CV - coefficient of variation
na – not applicable
PCB – polychlorinated biphenyl
TEQ – toxic equivalent
VCA – variance components analysis

The inference that can be made about the variance components in this dataset is limited due to data restrictions. An ideal design for assessing variance components is one that is balanced (i.e., equal replication at all levels) and, when variances are high (i.e., CVs close to 1), that has sufficient replication to minimize anomalous results. The variance estimates in this assessment of the clamming area composites were generated from the homogenization replicates run on one composite sample. One conclusion that

[7] Following homogenization of the 68 field samples, approximately equal volumes of sediment were transferred onto 2 stainless steel baking trays. A 30-square grid was created and equal aliquots of the homogenized sediment were collected from each grid square to fill the analytical sample jars. This process was repeated a total of 3 times from the same 30-square grids on the 2 trays to produce the triplicate samples.

can be cautiously drawn from the results shown in Table B2-3 is that large-scale spatial variance is high for both total PCBs and cPAHs. Large-scale spatial variance was the dominant variance source for the PCB results, whereas the cPAH results also had high within-composite variance (which included micro-scale spatial variance, homogenization, and analytical variance). Total PCBs had a high total CV of 128% (Table B2-3), with the within-composite variance estimated to be < 1% of the total, leaving > 99% of the total variance to large-scale spatial variability. The cPAHs had a high total CV of 116% (Table B2-3), with the within-composite variance estimated as 31% of the total, which is likely a reflection of the inherent variability of cPAHs because they are present in many different forms in the environment. Total organic carbon had a low total CV of 23%, and the within-composite and between-composite variances were approximately equal (Table B2-3), indicating that TOC was consistent on multiple spatial scales.

B2.2.2 Beach play areas

The 0–45-cm sediment dataset for beach play areas consisted of 24 composite samples (3 composites from each of the 8 beaches, the number of grab samples included in each composite ranging from 3 to 9, based on beach size) plus 6 field duplicates. All composite samples were analyzed for PCBs, arsenic, cPAHs, and dioxins/furans.

All four risk drivers had 100% DF, although the constituent compounds for dioxins/furans were not always detected. Total PCBs were calculated as the sum of detected Aroclors; at least one Aroclor was detected in each sample. The individual cPAH compounds were 100% detected in this dataset. For dioxins/furans, non-detected congeners contributed ≤ 1% of the dioxin/furan TEQ for most of the samples; in the remaining 6 samples, the TEQ contributions from non-detects ranged from 2 to 36%. The samples in which non-detects had the highest percent contribution to the total TEQ were those with the lower total TEQ values**.** The minor presence of non-detects in this dataset did not negatively affect the utility of these data to estimate site-wide mean and 95UCL estimates.

B2.2.2.1 95UCL calculations

Sediment DQO 9 (as applied to beach play area sediments) required that the 95UCL for the mean of each beach be established for the four risk drivers. The 95UCL for the baseline composite samples from each beach was calculated using Chebyshev's inequality (Equation 3). [8]

[8] The shape of the distribution could not be adequately evaluated with only three samples, so a non-parametric Chebyshev interval was used.

$$95UCL = \bar{X} + \sqrt{\left(1/0.05 - 1\right)} \times SE \qquad \textbf{Equation 3}$$

Where the additional term is defined as:

SE = standard error, or standard deviation of the mean, calculated from 3 beach-wide composites at each beach

The summary statistics for the four risk drivers in this dataset are presented in Table B2-4. The field duplicate samples were used to assess a combination of spatial variance within the sampling locations, homogenization variance, and analytical variance; they were not included in the calculation of the beach-wide means or standard errors (SEs) [9] for Beaches 1 and 6. This allowed for similar interpretations of the 95UCL estimates from every beach (i.e., each UCL represented the confidence limit for the mean of three composites).

[9] In the Pre-Design Studies database, parent and field duplicate results were retained as discrete samples (Windward and Integral 2017). The field duplicates were intended as quality assurance / quality control samples, to be used to evaluate the efficiency of field contamination procedures and the variability attributable to sample handling (sediment QAPP)—hence the decision to use only the parent sample results for baseline summaries (Windward 2018a).

Table B2-4. Summary statistics in beach play areas for intertidal (0–45-cm) sediments

Location	Total PCB Aroclors (µg/kg)			cPAH TEQ (µg/kg)			Dioxin/Furan TEQ (ng/kg)			Arsenic (mg/kg)		
	Mean	CV	95UCL[b]	Mean	CV	95UCL[b]	Mean	CV	95UCL[b]	Mean	CV	95UCL[b]
Beach 1	120	108%	445	169	101%	600	1.61	19%	2.38	14.7	63%	37.9
Beach 2	102	30%	179	276	60%	696	15.7	63%	40.7	44.7	25%	73.2
Beach 3	110	103%	396	100	90%	325	4.37	90%	14.3	4.01	23%	6.31
Beach 4	359	51%	815	45	42%	93.4	30.0	126%	125	6.24	36%	11.8
Beach 5	114	35%	214	1150	144%	5310	5.29	19%	7.87	8.74	40%	17.5
Beach 6	561	72%	1580	1343	9%	1650	13.2	56%	31.7	44.6	47%	96.8
Beach 7	65.2	58%	160	43	20%	63.4	2.13	10%	2.69	5.44	18%	7.97
Beach 8	123	58%	302	108	47%	232	4.05	28%	6.86	7.72	27%	13.0

[a] SE = SD/ sqrt(3).

[b] 95UCL calculated using Chebyshev's Inequality (n = 3 all areas). These estimates did not use the field duplicates taken at Beaches 1 and 6.

95UCL – 95% upper confidence limit (on the mean)
cPAH – carcinogenic polycyclic aromatic hydrocarbon
CV – coefficient of variation
PCB – polychlorinated biphenyl
SD – standard deviation
SE – standard error (of the mean)
TEQ – toxic equivalent

B2.2.2.2 Evaluation of sampling variance

The sampling variance is the total variance observed in the dataset and includes spatial, compositing, and analytical variance. Field and lab replicate results were used to quantify how much small-scale spatial heterogeneity and analytical variance contributed to the total variance in these data. This evaluation is useful for interpreting the current dataset, as well as providing information for modifying future sampling efforts.

Field duplicates were collected at Beaches 1 and 6. At each sampling location in these two areas, sediment from each hole was placed in two 16-oz jars (rather than one) for the field duplicates. The field duplicates were composited following the same methods and using the same locations as the original beach composite samples. A VCA (Section 1.3) quantified small-scale spatial variance (i.e., the differences among duplicate field samples at both beaches) and laboratory variance (for cPAHs only, the differences among laboratory triplicates in one composite sample) relative to the total variance within a beach (Table B2-5). For PCB Aroclors, the variability among field duplicates was low (≤ 6%), indicating relative consistency among grab samples taken from the same holes. For the other analytes, small-scale spatial variability was 95% or more of the total variance for the two beaches, with one exception: Arsenic at Beach 6 had small-scale spatial variability that contributed 33% to the total. Laboratory variance was evaluated for only cPAHs at Beach 1, and it was effectively 0% of the total.

Table B2-5. Results of VCA for beach play area composites.

Variance Source	Degrees of Freedom	Sum of Squared Error	Mean Squared Error	Variance Component	% Total for Variance Components
Total PCBs					
Beach 1 (CV = 109%)[a]					
Total observed	2.1	na	na	23,225	100
Among composite locations	2	90,174	45,087	21,862	94
Within composite locations[b]	3	4,090	1,363	1,363	6
Beach 6 (CV = 67%)					
Total observed	2.0	na	na	139,563	100
Among composite locations	2	555,072	277,536	137,973	99
Within composite locations[b]	3	4,769	1,590	1,590	1
cPAH TEQ					
Beach 1 (CV = 138%)					
Total observed	4.6	na	na	222,291	100
Among composite locations	2	603,289	301,644	0[c]	0[c]
Within composite locations[b]	3	777,622	259,207	221,498	> 99
Laboratory[d]	2	1,586	793	793	< 1
Beach 6 (CV = 136%)					
Total observed	4.7	10,327,609	na	na	100

Variance Source	Degrees of Freedom	Sum of Squared Error	Mean Squared Error	Variance Component	% Total for Variance Components
Among composite locations	2	21,589,026	10,794,513	466,904	5
Within composite locations[b]	3	29,582,115	9,860,705	9,860,705	95
Dioxins/furans TEQ					
Beach 1 (CV = 44%)					
Total observed	4.8	na	na	0.8	100
Among composite locations	2	1.2	0.6	0[c]	0[c]
Within composite locations[b]	3	2.4	0.8	0.8	100
Beach 6 (CV = 57%)					
Total observed	4.8	na	na	88	100
Among composite locations	2	87	43	0[c]	0[c]
Within composite locations[b]	3	264	88	88	100
Arsenic					
Beach 1 (CV = 52%)					
Total observed	4.8	na	na	64	100
Among composite locations	2	78	39	0[c]	0[c]
Within composite locations[b]	3	191	64	64	100
Beach 6 (CV = 47%)					
Total observed	2.8	na	na	360	100
Among composite locations	2	1,205	602	242	67
Within composite locations[b]	3	353	118	118	33

[a] CV among all 6 composite samples.

[b] The variability among three pairs of field duplicates composed of sediment taken from the same holes as the parent samples.

[c] Negative variance component estimate set to 0 (see Section 1.3).

[d] The variability among laboratory triplicates of a single composite sample (LDW18-IT45-B1-Comp1).

cPAH – carcinogenic polycyclic aromatic hydrocarbon
CV – coefficient of variation
na – not applicable
PCB – polychlorinated biphenyl
TEQ – toxic equivalent
VCA – variance components analysis

The inference that can be made about the variance components in this dataset is limited due to data restrictions. An ideal design for assessing variance components is one that is balanced (i.e., equal replication at all levels assessed) and, when variances are high (i.e., CVs close to 1), has sufficient replication to limit anomalous results. The variance estimates for small-scale spatial variability in this assessment are balanced. However, the estimate of laboratory variability of cPAHs in Beach 1 was restricted to the variance observed within a single sample. Conclusions that can be cautiously drawn from the results shown in Table B2-5 are that small-scale spatial variability comprises most of the total variance for dioxins/furans, and, at least at Beach 1, for arsenic, but that the total variance is relatively low (CVs ≤ 57%, Table B2-5). The total variance for cPAHs is high, with CVs ≥ 136% for the two beaches, and most of this variance is small-scale spatial variability (> 95%, Table B2-5).

The influence of small-scale spatial variability (variance within composite locations) on the estimated results was evaluated by comparing the mean and 95UCL results both with and without the field duplicates at Beaches 1 and 6 (Table B2-6). The calculations excluding the field duplicates used only the three primary samples and Chebyshev's inequality for the 95UCL with two degrees of freedom. The calculations including the field duplicates used all six results, first averaging the two duplicates for each composite, and then calculating the 95UCL with two degrees of freedom using Chebyshev's inequality. Consistent with the VCA, when small-scale spatial variability was found to be a small percentage of the total (PCBs at both beaches and arsenic at Beach 6), the difference between the two 95UCLs was minimal. The most variable results for 95UCLs with and without field duplicates were observed for cPAH TEQs. This variability was not widespread: Field duplicates at Beach 1 had high variance between duplicates from only one of the composites (LDW18-IT45-B1-Comp1), while at Beach 6 the variance was high between all three pairs of field duplicates.

Table B2-6. Effect of field duplicates on means and 95UCLs at Beaches 1 and 6

COC	Statistic	Beach 1		Beach 6	
		No Field Duplicate[a] (n=3, df=2)	With Field Duplicate (n=6, df=2)	No Field Duplicate (n=3, df=2)	With Field Duplicate (n=6, df=2)
Arsenic (mg/kg)	Mean	14.7	15.32	44.6	40.2
	SE	5.31	2.54	12.0	10.0
	95UCL[b]	37.9	26.4	96.8	83.9
Total PCB Aroclors (ug/kg)	Mean	120	140	561	554
	SE	74.5	86.7	234	215
	95 UCL	445	518	1582	1491
cPAH TEQ (ug/kg)	Mean	169	336	1343	2368
	SE	98.7	268	71.3	1341
	95UCL	600	1504	1654	8214
Dioxin/furan TEQ (ng/kg)	Mean	1.61	2.04	13.2	16.5
	SE	0.178	0.318	4.23	2.69
	95UCL	2.38	3.42	31.7	28.3

[a] When field duplicates were excluded, only the three composite samples were used. When field duplicates were included, the mean of the two composite samples with each sample ID (sample, and sample-FD) were averaged, and summary statistics were calculated from the three means per beach.

[b] As indicated in the surface sediment QAPP (Windward 2018a), the 95UCL for beaches were calculated using Chebyshev's Inequality (n = 3 all areas).

95UCL – 95% upper confidence limit (on the mean)
COC – contaminant of concern
cPAH – carcinogenic polycyclic aromatic hydrocarbon
ID – identification
PCB – polychlorinated biphenyl
QAPP – quality assurance project plan
SE – standard error (of the mean)
TEQ – toxic equivalent

The RMEs for the beach play area sediment datasets were high, partially due to the conservatively high Chebyshev 95UCL estimate, which was required because of the

small sample size. In the context of the project objectives, smaller RMEs are not needed for total PCBs in beach play sediments because the RBTC was met at all beaches.

Pre-Design Studies results indicate that concentrations of cPAHs and dioxins/furans exceed the cleanup values at several beaches, resulting in the need for design sampling at these beaches. Following any remediation, mean and variances will be lower; sufficiency of the sampling design can be evaluated at that time.

B3 Surface Water

This section provides statistical details regarding the interpretation of the surface water data, as presented in Section 3 of the main report. Surface water grab samples were summarized in the main report, and no further discussion of these data is needed in this appendix. However, the C_{free} of total PCBs from the passive samplers are discussed herein with respect to mean, variance, and distribution.

B3.1 Distribution of Passive Sampler Results

During the development of the *Pre-Design Studies Work Plan* (Windward and Integral 2017), hereafter referred to as the Work Plan, the passive sampler study design was developed using the most recent passive sampler data from the LDW (i.e., passive sampler data from Apell and Gschwend 2017). The Apell and Gschwend (2017) passive sampler data were limited to a single sample at three different locations. These data were insufficient to adequately evaluate the distributional form. Consequently, the *a priori* power calculations for the Work Plan were based on untested assumptions about the distributional form of the data and used both the normal and log-normal distributions. The total PCB baseline dataset was sufficient (n = 35)[10] to investigate the distributional form, so it was evaluated graphically using normal probability plots and formally using goodness-of-fit tests (Section 1.1). The normal probability plot for the station residuals[11] (Figure B3-1) indicated that data were approximately normally distributed; the Shapiro-Wilk goodness-of-fit test did not reject normality (p=0.60). Consequently, a parametric ANOVA model may be used to assess these data.

[10] Nine passive sampler replicates were deployed at each location in both baseline years (total n = 36). One passive sampler result was rejected from location PS1 in 2018 (Windward [in prep]), resulting in a total n = 35 for the baseline passive sampler dataset.

[11] Residuals are the individual observations minus the station mean. The station residuals have a common mean (zero), which allows results from the two stations to be pooled to evaluate the shape and variance of these data.

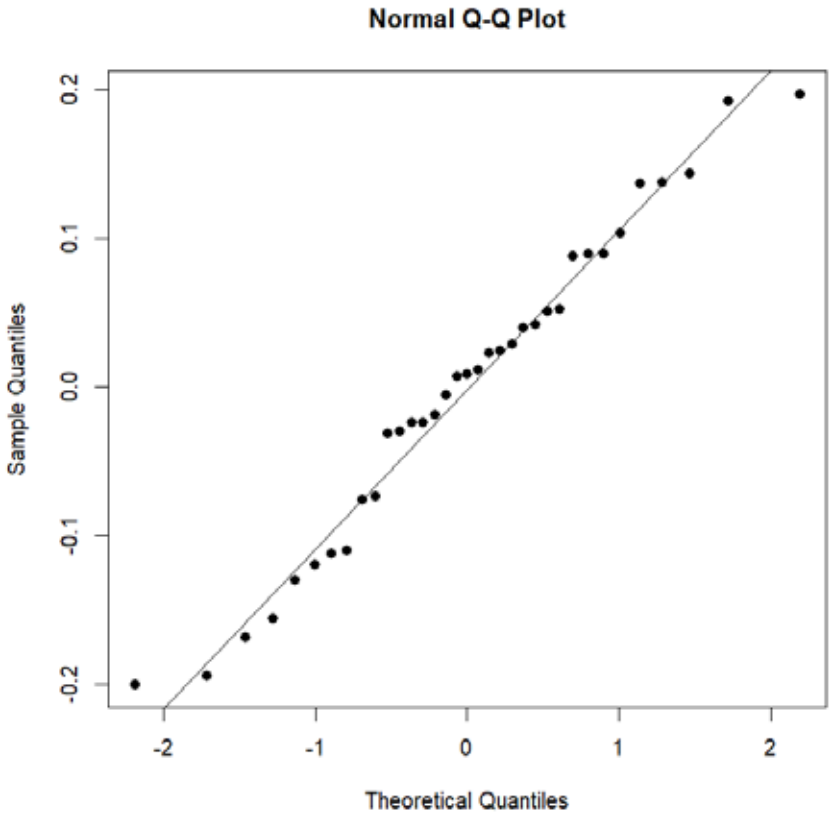


Figure B3-1. Normal probability plot of station residuals for the baseline passive sampler dataset (n=35)

B3.2 EVALUATION OF SAMPLING VARIANCE

This passive sampler dataset provides a nearly balanced sampling design, with eight or nine replicates at both locations in both years. The statistical significance of Location and Year effects was tested using a two-factor ANOVA model, including an interaction term (Location × Year) that allowed for the possibility that the two locations did not respond similarly over time (Table B3-1). The interaction term was not statistically significant ($p = 0.26$), indicating that any differences observed between the two stations were similar for each of the two baseline years. "Location" in Table B3-1 summarizes the differences between stations, averaged over the two years; the average difference between stations (29 pg/L) was not statistically significant ($p = 0.45$). "Year" in Table B3-1 summarizes the differences between years, averaged over the two stations; the average difference in years (265 pg/L) was highly statistically significant ($p < 0.001$).

Table B3-1. ANOVA table for comparison of total PCBs in passive samplers between two locations and two baseline years (2017 and 2018)

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic[a]	p-Value
Year	1	624,280	624,280	53.113	3.36E-08
Location	1	6,953	6,953	0.592	0.448
Location × year	1	15,477	15,477	1.317	0.26
Residuals	31	364,369	11,754	-	-

The F statistic is the ratio of appropriate mean squares, which is used to assess the significance of the source of variance that is indicated by the p-value.

ANOVA – analysis of variance

PCB – polychlorinated biphenyl

Using a VCA (Section 1.3), a relative comparison of the variance among location and year to the residual variability was made. The variance components for the total PCB passive sampler dataset are summarized in Table B3-2. The variability between locations was effectively 0%, and the variability among replicate samplers was 25%. Most of the variability (i.e., 74% of the total) was between years.

Table B3-2. Results of VCA for total PCB passive sampler data

Variance Source	Degrees of Freedom	Sum of Squared Error	Mean Squared Error	Variance Component	% Total for Variance Components
Total observed	1.7	na	na	0.047	100%
Location	1	0.011	0.011	0[a]	0[a]
Year	1	0.620	0.620	0.035	74%
Location × year	1	0.015	0.015	0.000	1%
Residual	31	0.364	0.012	0.012	25%

Note: Design is slightly unbalanced with 8 replicates in one Location-Year combination, and 9 replicates in the other three combinations.

[a] Negative variance component set to zero (see Section 1.3).

na – not applicable

PCB – polychlorinated biphenyl

VCA – variance components analysis

From these results, it is possible to infer that location contributes very little to the variability in the dataset. The two passive sampler locations were selected to provide spatial coverage of the LDW (i.e., one location further downstream at RM 1.9 and one location further upstream at RM 3.3). In contrast, the annual differences constituted a large percentage of the variability of this dataset (74%, Table B3-2). The residual variability among the replicate samplers was relatively low (25%) and, most importantly, was very consistent between years and locations, suggesting that these samplers had high precision. From the evaluation of these data, temporal variability is greater than spatial variability within the LDW. The variability across locations accounts for essentially 0% of the total variance observed, which indicates that the two locations are redundant.

B3.3 POWER AND SAMPLE SIZE

The statistical approach for comparing the passive sampler data between baseline and future monitoring events is expected to be on a station-by-station basis. The variance estimate used in the *a priori* power analysis during Work Plan (Windward and Integral 2017) development was derived from the most recent passive sampler results from the LDW at that time (i.e., single observations from each of three locations (Apell and Gschwend 2017)). Using residual variability from the recent baseline dataset, the power analysis was updated to assess the expected minimum detectable difference (MDD) between baseline and future monitoring events. The variance estimates from the recent baseline dataset are summarized in Table B3-3.

Table B3-3. Summary statistics for sum of PCB congeners from PE samplers deployed in the LDW

Summary Statistic	Pre-Design Studies Dataset	
	2017	2018
Sample size	18 (9 reps per station)	17 (8 at PS1 and 9 at PS2)
C_{free} total PCB mean concentration ($\bar{x}$) (ng/L)	1.26 (1.25 at PS1 and 1.26 at PS2)	0.99 (1.03 at PS1 and 0.96 at PS2)
SD for C_{free} total PCBs (ng/L)	0.115[a] (0.101 at PS1 and 0.128 at PS2)	0.101[a] (0.115 at PS1 and 0.086 at PS2)
CV = SD / $\bar{x}$	9.2%[b]	10.1%[b]

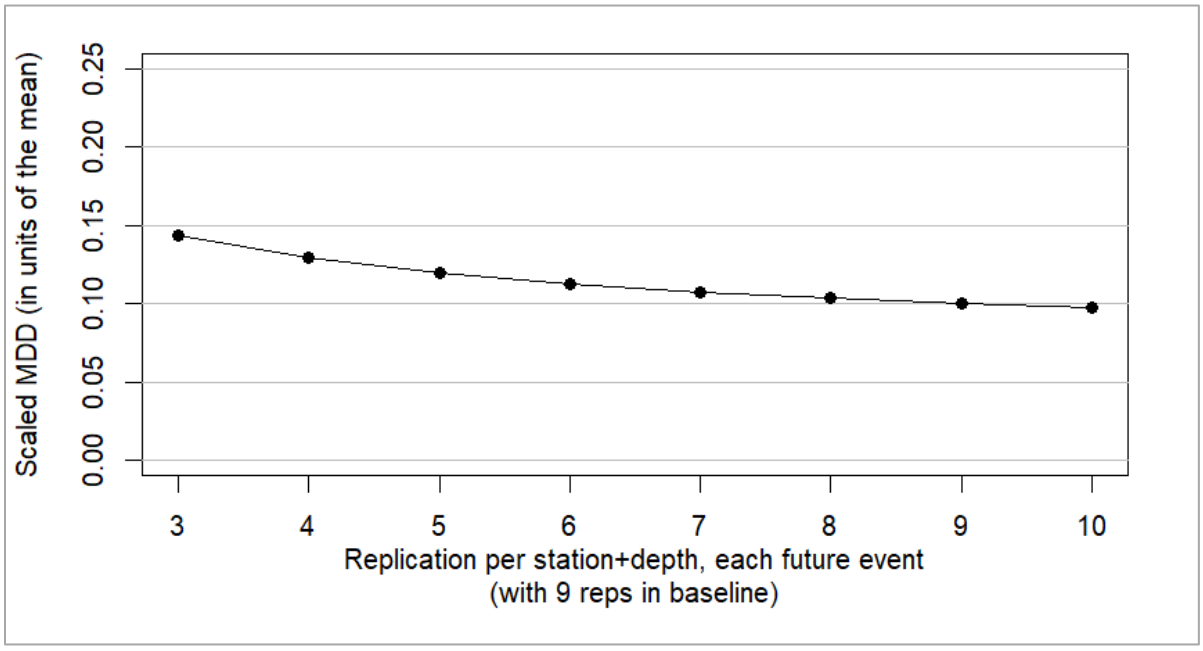
[a] The combined SD values reported for the Pre-Design Studies baseline samples are the residual standard errors across both stations within each sampling year.

[b] The CVs reported for Pre-Design Studies baseline data use the values combined across the two stations.

CV – coefficient of variation
LDW – Lower Duwamish Waterway
PCB – polychlorinated biphenyl
PE – polyethylene
SD – standard deviation
SE – standard error (of the mean)

The standard deviations (SDs) within the Pre-Design Studies passive samplers were very similar for the two baseline years (0.1 ng/L), and were low relative to the mean (CV of < 10%). Replicate variability was low for these passive samplers due to a combination of the extended exposure period (which avoided variance induced from short-term temporal fluctuations) and low instrument measurement error. Using a mean concentration of total PCBs for the Pre-Design Studies samples (2017–2018) of 1.125 ng/L, and a CV of approximately 10%, the MDD for a comparison between baseline and future monitoring is approximately 0.1125 ng/L for nine replicates (Type I and Type II errors = 0.10; comparison using a nested ANOVA model with two years nested within each study period [baseline and future] at a single location). Reducing replication in future monitoring events to as few as three would be expected to result in very tight MDDs of less than 15% of the mean (Figure B3-2). Five passive sampler results in future years would allow sufficient replicates to confirm the normality of the data and still achieve a low MDD (approximately 12%) for comparisons to baseline; a

total of nine samplers could still be deployed to ensure sufficient data in the event that samplers were lost in the field or analytical or variance issues were raised.



Note: Assumes a parametric 2-tailed *t*-interval testing for the difference of means between baseline (2 years) and future (2 years). Types I and II errors are both set at 10%. The CV value of 0.10 was observed during baseline sampling. Uses 9 replicates per station+depth during baseline sampling event. MDD (on the y-axis) is expressed as a percent of the baseline mean.

Figure B3-2. Relationship between replication within each station/depth for future sampling event and scaled MDD

B4 Fish and Crab Tissue

This section provides statistical details regarding the interpretation of the fish and crab tissue data, as presented in Section 4 of the main report.

B4.1 Influence of Non-Detects

Within the baseline fish and crab tissue datasets, data below detection had a noticeable influence only on the dioxin/furan results for the graceful crab edible meat samples. These non-detected dioxin/furan compounds introduced uncertainty to the calculated TEQ and affected whether the sample result was above or below the TTL in 6 of the 12 samples. Six graceful crab edible meat samples that had a dioxin/furan TEQ below the TTL using ½ the reporting limit (RL) for the non-detected compounds would have TEQs greater than the TTL if the full RL was used instead (with a TTL exceedance of 15% or less).

None of the other datasets were notably affected by non-detects. Total PCBs were calculated as the sum of detected Aroclors. Individual cPAH compounds were not detected in any of the crab tissue samples, and cPAHs were not analyzed in fish tissues because of the ability of fish to metabolize cPAH compounds. For dioxins/furans in

tissues for which target tissue levels (TTLs) were available (i.e., graceful crab whole body and edible meat, and English sole whole body), the non-detected compounds notably affected only the graceful crab edible meat samples as noted above.

B4.2 95UCL CALCULATIONS

Fish and crab tissue DQO 1 requires that a 95UCL for the site-wide mean be calculated for each tissue type. The sampling approach used a stratified design to account for possible differences of mean and variability in composite tissue concentrations across reaches and subreaches. As appropriate for the stratified sampling design, residuals[12] within each reach or subreach were combined, and methods (described in Section 1.1) were used to identify the best distributional form for each COC and tissue type. The normal distribution was preferred for this stratified model; in all cases, the normal distribution provided a reasonable fit to the data, and no outliers were present (Table B4-1 and Figures B4-1 and B4-2). Total PCB Aroclors in several tissue datasets (i.e., English sole whole body [calculated], shiner surfperch, and graceful crab whole body [calculated]) showed some deviations relative to the normal distribution. However, these datasets passed the goodness-of-fit tests for normality, and the general symmetry and lack of extreme values within these datasets indicate acceptability of the normal distribution for calculating the 95UCLs.

Table B4-1. Goodness-of-fit and relative variance summary for COCs in baseline fish and crab tissues

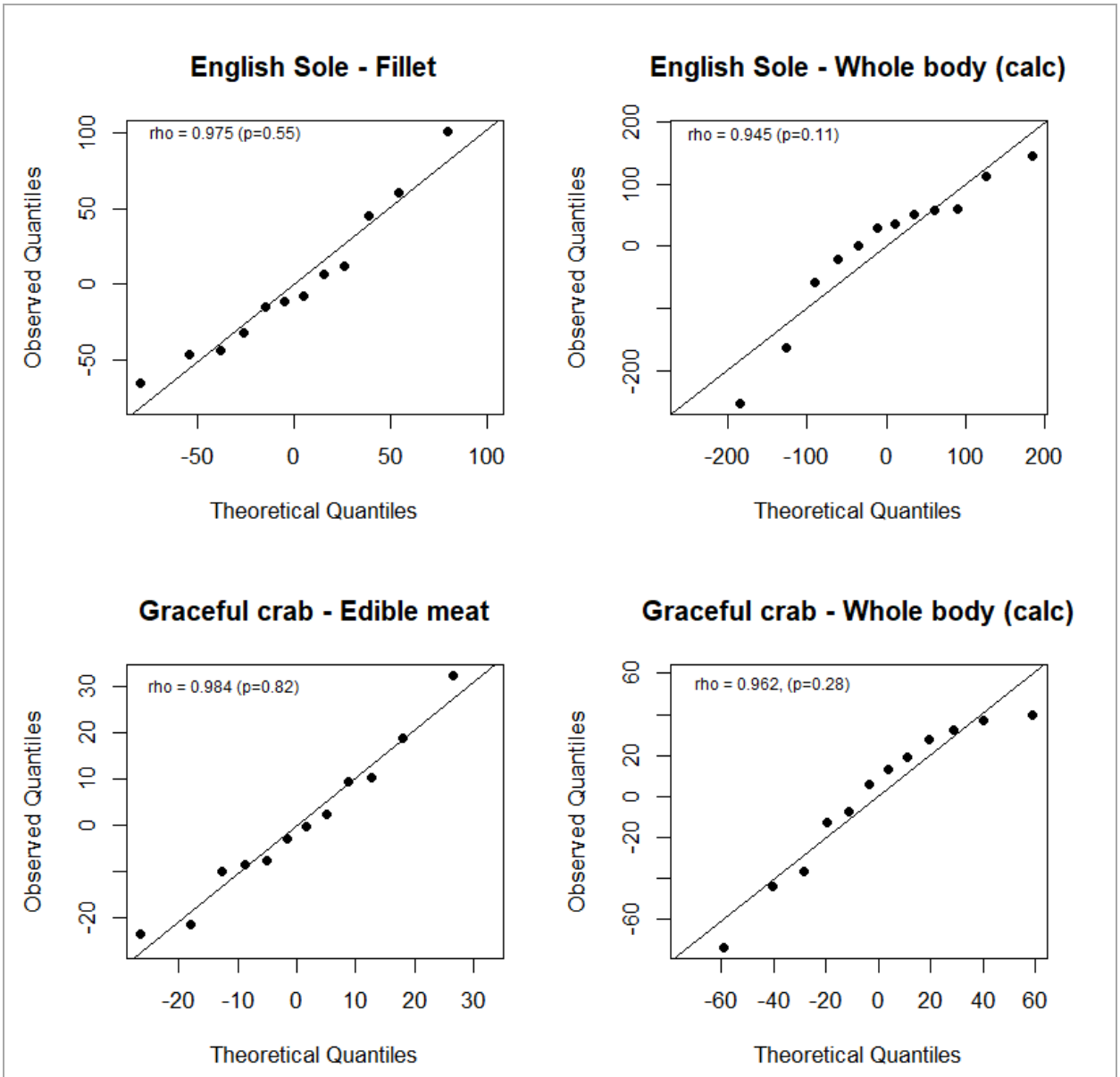
COC	Species and Tissue Type	Normal PPCC[a]	p-Value[b]	CV
Total PCBs (sum of Aroclors) (ug/kg ww)	English sole – fillet	0.975	0.55	0.20
	English sole – whole body (calculated)	0.945	0.11	0.16
	graceful crab – edible meat	0.984	0.82	0.15
	graceful crab – whole body	0.962	0.28	0.15
	shiner surfperch – whole body	0.956	0.20	0.08
Dioxin/furan TEQ (ng/kg ww)	English sole - fillet	0.988	0.91	0.20
	English sole – whole body (calculated)	0.978	0.64	0.11
	graceful crab – edible meat	0.964	0.32	0.19
	graceful crab – whole body	0.986	0.87	0.16
	shiner surfperch – whole body	0.984	0.83	0.28

[a] PPCC for the normal distribution.

[b] p-value for the PPCC goodness-of-fit test.

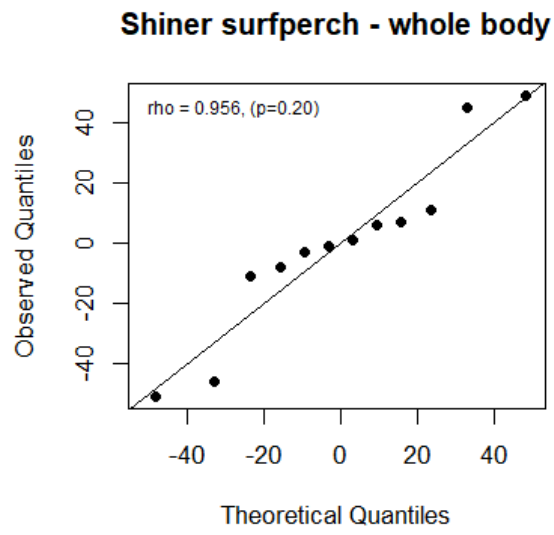
COC – contaminant of concern
CV – coefficient of variation
PCB – polychlorinated biphenyl
PPCC – probability plot correlation coefficient
TEQ – toxic equivalent
TTL – target tissue level
ww – wet weight

[12] Goodness-of-fit was applied to the residuals from a stratified model (i.e., the differences between each composite value and the mean for all samples from the same LDW river reach).



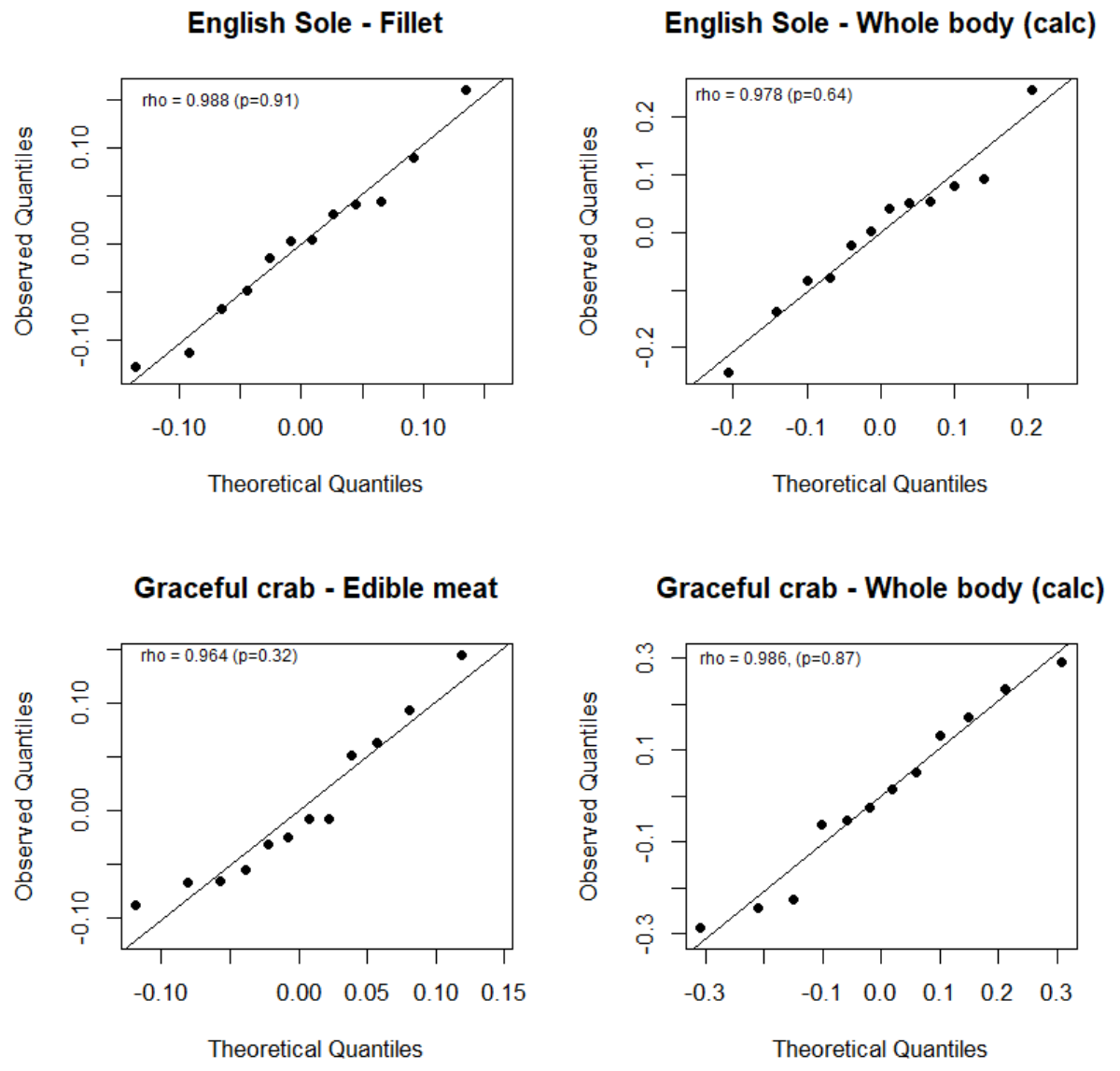
Results are shown for combined residuals from each LDW river reach, as appropriate for the stratified model.

Figure B4-1a. Normal probability plots of baseline total PCB Aroclors (µg/kg ww) in fish and crab composite tissue samples



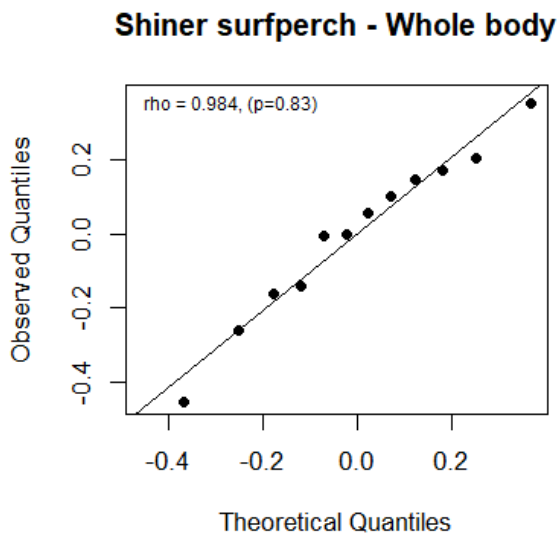
Results are shown for combined residuals from each LDW river reach, as appropriate for the stratified model.

Figure B4-1b. Normal probability plots of baseline total PCB Aroclors (µg/kg ww) in fish and crab composite tissue samples



Note: Results are shown for combined residuals from each LDW river reach, as appropriate for the stratified model.

Figure B4-2a. Normal probability plots of baseline dioxin/furan TEQ (ng/kg ww) in fish and crab composite tissue samples



Note: Results are shown for combined residuals from each LDW river reach, as appropriate for the stratified model.

Figure B4-2b. Normal probability plots of baseline dioxin/furan TEQ (ng/kg ww) in fish and crab composite tissue samples

B4.3 Statistical Comparisons Between Concentrations of Total PCBs in 2017 Samples (Pre-Design Studies Dataset) and 2007 Samples

The temporal change in tissue concentrations between 2007 and 2017 was evaluated, to the extent allowed by the data, with statistical comparisons made between the average tissue concentrations of samples collected in 2007 as part of the LDW remedial investigation (RI) (Windward 2010) and the average tissue concentrations of samples collected as part of the Pre-Design Studies in 2017. The designs from the two studies were generally similar, and any differences (primarily the number of fish/crab per composite) are noted in the following sections. In some cases, the data that were available precluded making site-wide comparisons, and statistical comparisons could only be made on a smaller spatial scale than the entire LDW.

Parametric ANOVA models were fit to each dataset and standard residual diagnostic plots (i.e., normal probability plots, residuals vs. leverage values, and fitted values vs. standardized and unstandardized residuals) were used to assess the appropriateness of each model. If there were issues apparent in the diagnostic plots regarding the assumptions for the parametric ANOVA model, the statistical results were caveated.

In this section, multiple statistical comparisons are made, and with a Type I error (α) of 0.05, 5% of these comparisons may be statistically significant purely by chance—which is what the Type I error rate represents. The comparison-wise p-values were reported in the main text to indicate the strength of difference (or lack thereof) between the two

study efforts. These results are meant to be informative of a pattern, not necessarily definitive statements regarding statistical significance.

B4.3.1 English sole fillet and whole-body tissues

Data for total PCB Aroclors in fillet and whole-body English sole samples for 2007 and 2017 are shown in Table B4-2. The sampling designs for the 2007 and 2017 samples were comparable—similar sampling areas were used and the sizes of the fish included in the composites were similar. However, the number of individual fish per composite was 5 in 2007 and 10 in 2017. The study design had sampling reach nested within year, and replication within reach for each of two years. These data were analyzed with an ANOVA model for a nested design (Table B4-3).

Table B4-2. Summary of total PCB results in English sole fillet and whole-body tissues for baseline and RI datasets

Dataset	English Sole Fillet		English Sole Whole Body			
	Average Total PCB Aroclor Concentration (µg/kg ww)	Count	Average Total PCB Aroclor Concentration (µg/kg ww)	Count	Average Total PCB Congener Concentration (µg/kg ww)	Count
LDW RI 2007						
Reach 1	318	6	609	12	1,290	4
Reach 2	403	3	809[a]	7	1,980	2
Mean of Reaches 1 and 2	361	9	709	19	1,640	6
LDW Baseline 2017						
Reach 1	342	6	888	6	1,010	3
Reach 2	177	6	621	6	606	3
Mean of Reaches 1 and 2	259	12	704	12	808	6

[a] An extreme concentration was identified in this reach in 2007 (1,600 µg/kg ww). When excluded, the reach mean concentration was 677 µg/kg ww, and the 2007 mean was 643 µg/kg ww. The influence of this sample was evaluated by comparing the two years without this sample. Tables B4-5 and B4-3 show the ANOVA model results.

ANOVA – analysis of variance
LDW – Lower Duwamish Waterway
PCB – polychlorinated biphenyl
RI – remedial investigation
ww – wet weight

Table B4-3. ANOVA table for comparison of total PCBs in English sole tissue between 2007 and 2017

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic	p-Value
Total PCB Aroclors in English sole fillet (2007 vs. 2017)					
Year	1	39,353	39,353	6.581	0.02007
Reach within year	2	95,089	47,544	7.951	0.00365
Residuals	17	101,655	5,980	-	-
Total PCB Aroclors in English sole whole body (2007 vs. 2017)					
Year	1	37,900	37,900	0.645	0.4291
Reach within year	2	390,995	195,498	3.325	0.0512
Residuals	27	1,587,496	58,796	-	-
Total PCB Aroclors in English sole whole body (2007 vs. 2017), excluding outlier from Reach 2 in 2007					
Year	1	108,486	108,486	3.292	0.0812
Reach within year	2	233,429	116,715	3.542	0.0436
Residuals	26	856,744	32,952	-	-
Total PCB congeners in English sole whole body (2007 vs. 2017)[a]					
Year	1	1.53E+9	1.53E+9	5.18	0.0524
Reach within year	2	8.74E+8	4.37E+8	1.479	0.284
Residuals	8	2.36E+9	2.95E+8	-	-

[a] Normality and homogeneity of variances assumptions were challenged by the 2007 data from Reach 2. These results are approximate.

ANOVA – analysis of variance PCB – polychlorinated biphenyl

B4.3.2 Shiner surfperch

Data for total PCB Aroclors in whole-body shiner surfperch samples for 2007 and 2017 are shown in Table B4-4. The sampling designs for the 2007 and 2017 samples were comparable—similar sampling areas were used, although the areas sampled during baseline were larger, and the sizes of the fish included in the composites were similar. However, the number of individuals per composite was 10 in 2007 and 15 in 2017. The study design was nested, with replication within subreach for each of two years. These data were analyzed with an ANOVA model for a nested design (Table B4-5).

Table B4-4. Summary of total PCB results in shiner surfperch whole-body tissues for 2007 and 2017 datasets

Dataset	Average Total PCB Aroclors Concentration (µg/kg ww)	Count	Average Total PCB Congener Concentration (µg/kg ww)	Count
LDW RI 2007				
Subreach 1a (T1)	268	6	739	2
Subreach 1b (T2)	415	6	525	2
Subreach 2a (T3)	763	6	1,783	2
Subreach 2b (T4)	315	4	--	0

Table B4-4. Summary of total PCB results in shiner surfperch whole-body tissues for 2007 and 2017 datasets

Dataset	Average Total PCB Aroclors Concentration (µg/kg ww)	Count	Average Total PCB Congener Concentration (µg/kg ww)	Count
All subreaches combined	440	22	1,016	6
LDW Baseline 2017				
Subreach 1a	439	3	496	2
Subreach 1b	370	3	405	2
Subreach 2a	504	3	551	2
Subreach 2b	316	3	333	2
All subreaches combined	407	12	446	8

LDW – Lower Duwamish Waterway
PCB – polychlorinated biphenyl
ww – wet weight

Table B4-5. ANOVA table for comparison of total PCB Aroclor data in shiner surfperch samples

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic	p-Value
Total PCB Aroclors in shiner surfperch whole body (2007 vs. 2017)[a]					
Year	1	15,481	15,481	0.672	0.420
Reach within Year	6	927,316	154,553	6.713	0.0002
Residuals	26	598,611	23,024	-	-
Total PCB congeners in shiner surfperch whole body (2007 vs. 2017), Reach 1 only[a]					
Year	1	37,900	37,900	0.645	0.4291
Reach within Year	2	390,995	195,498	3.325	0.0512
Residuals	27	1,587,496	58,796	-	-

[a] Normality and homogeneity of variances assumptions were challenged by the 2007 data. These results are approximate.

ANOVA – analysis of variance
PCB – polychlorinated biphenyl

B4.3.3 Graceful crab

Data for total PCB Aroclors in graceful crab tissue samples for 2007 and 2017 are shown in Table B4-6. Insufficient Dungeness crab samples were available for these two sampling years to conduct a statistical comparison, so only graceful crab was evaluated.[13] The sampling designs for the 2007 and 2017 samples were comparable—similar sampling areas were used, and the sizes of the crabs included in the composites were similar. However, the number of individuals per composite was 5 crabs in 2007

[13] Graceful crab is more commonly available in the LDW than Dungeness crab, so graceful crab was collected for the purpose of trend evaluations.

and 7 to 14[14] crabs in 2017. No graceful crab samples were available from Reach 2 in 2007, so the statistical comparisons between years were conducted using only samples from Reach 1. These data were analyzed using a single-factor ANOVA model to test for differences between the years (Table B4-7).

Table B4-6. Summary of total PCB Aroclor results in graceful crab tissues for the 2007 and 2017 datasets

Dataset	Edible Meat		Whole Body	
	Average Total PCB Concentration (µg/kg ww)	Count	Average Total PCB Concentration (µg/kg ww)	Count
LDW RI 2007				
Reach 1 (T1 & T2)	41	6	155	6
Reach 2 (T3 & T4)	na	0	na	0
Reaches 1 and 2 combined	41	6	155	6
LDW Baseline 2017				
Reach 1	146	6	319	6
Reach 2	84	6	192	6
Reaches 1 and 2 combined	115	12	255	12

LDW – Lower Duwamish Waterway
na – not available
PCB – polychlorinated biphenyl
RI – remedial investigation
ww – wet weight

Table B4-7. ANOVA table for comparison of total PCB Aroclor data in graceful crab samples (Reach 1 only)

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic	p-Value
Graceful crab edible meat (Reach 1 only)					
Year	1	33,444	33,444	231.9	3.03E-08
Residuals	10	1,442	144	-	-
Graceful crab whole body (Reach 1 only)					
Year	1	80,688	80,688	31.79	0.000216
Residuals	10	25,384	2,538	-	-

ANOVA – analysis of variance
PCB – polychlorinated biphenyl

B4.4 Relationship Between PCB Aroclors and Congeners in Fish and Crab Tissues

A subset of the fish and crab tissue samples were evaluated for both PCB Aroclors and congeners (Table B4-8). Within each species and tissue type, the correlation between the two PCB estimates was evaluated for consistency of results. The slopes of the linear regressions between the Aroclor and congener sums were significantly different from

[14] Seven crab were included in the graceful crab edible meat composites, and 14 graceful crab were included in the hepatopancreas composites.

zero (adjusted p-value < 0.05, Table B4-8) for all tissues except English sole fillet. In addition, the regression slopes were all less than one, indicating that Aroclors under-predicted congeners. The slopes for the graceful crab tissues were very close to one, indicating very good consistency between the two total PCB estimates in these tissues. The data with the ordinary least squares regression lines are shown in Figure B4-3.

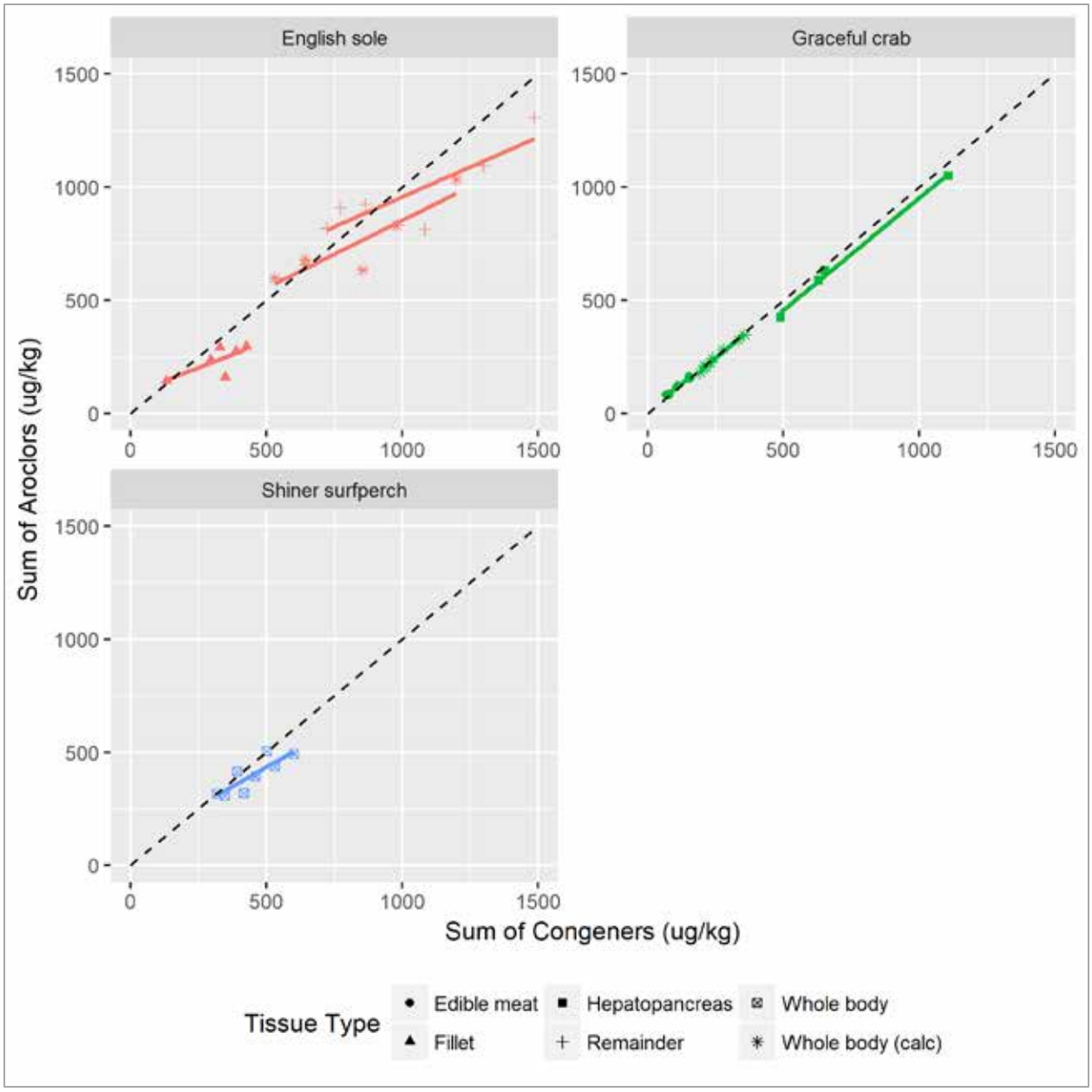
Table B4-8. Number and type of fish and crab tissue samples with PCB Aroclors and Congeners

Species - Tissue Type	No. of Samples with Both PCB Aroclors and Congeners	Regression R2	Slope of OLS Regression Line	Adjusted p-Value for Linear Slope[a]
English sole - fillet	6	0.51	0.46	0.11
English sole - remainder	6	0.72	0.53[b]	0.04
English sole - whole body (calculated)	6	0.82	0.60[b]	0.02
Graceful crab - edible meat	8	0.97	0.93[b]	0.00
Graceful crab - hepatopancreas	4	0.99	0.99[b]	0.01
Graceful crab - whole body (calculated)	8	0.98	0.99[b]	0.00
Shiner surfperch - whole body	8	0.71	0.68[b]	0.01

[a] The p-values were adjusted to control the false discovery rate, or Type I error rate, among rejected hypotheses for the set of regressions on 8 tissue types. The adjusted p-values were calculated using the *p.adjust(method="BH")* function in R (R Core Team 2018).

[b] Indicates regression slope is significantly different from 0 (i.e., adjusted p-value< 0.05).

OLS – ordinary least squares



Note: Each species is shown in its own panel, and different symbols indicate the tissue type. The OLS linear regression lines are shown for each species-tissue type combination (see Table B4-8 for regression results). The black dashed line on each panel is the 1:1 line.

Figure B4-3. Plot of Aroclors vs. congeners for baseline fish and crab tissues

B4.5 POWER AND SAMPLE SIZE

The relative variance estimates (CVs) used in the *a priori* power analysis during the development of the sampling design in the Work Plan (Windward and Integral 2017) were based on the RI tissue datasets. As a conservative assumption, these CVs were rounded up for use in the Work Plan. In an effort to reduce variance among samples and avoid outlier concentrations, each Pre-Design Studies composite sample was comprised of a larger number of individuals than was used in the RI tissue dataset.

Using the updated CV results from the Pre-Design Studies, the expected MDDs for comparison between baseline and future monitoring were calculated for total PCB Aroclors and dioxin/furan TEQ for each species and tissue type with TTLs (Table B4-9). For all tissue types, the CVs were better than the estimates used during Work Plan development. A consequence of a lower CV is increased statistical power for comparisons between baseline and future monitoring. The estimated MDDs ranged from 10 to 25% of the baseline mean, indicating that the baseline design is statistically sufficient to detect meaningful changes in tissue concentrations.

Table B4-9. Expected MDDs for comparisons between baseline and future monitoring for COCs and species/tissue types with TTLs

Chemical	Species and Tissue Type	Work Plan CV	Pre-Design Studies Site-wide Mean	Pre-Design Studies CV	MDD[a]	
					Conc.	% Pre-Design Studies Mean
Total PCBs (Aroclors) (µg/kg ww)	English sole – fillet	0.4	259	0.20	65	25%
	shiner surfperch – whole body	0.4	407	0.08	41	10%
	graceful crab – edible meat	0.25	115	0.15	21	19%
	graceful crab – whole body	0.25	255	0.15	48	19%
Dioxin/furan TEQ (ng/kg ww)	English sole – whole body	0.4	1.18	0.11	0.16	14%
	graceful crab – edible meat	0.25	0.406	0.19	0.10	24%
	graceful crab – whole body	0.25	1.21	0.16	0.24	20%

[a] The MDD is the minimum detectable difference for a comparison between baseline and a future monitoring event using the baseline study design. It assumes a nested ANOVA model, sampling identical to Pre-Design Studies, with sampling reach nested within year and total n = 12 in each year, Type I and Type II error rates set at 0.10.

ANOVA – analysis of variance
COC – contaminant of concern
CV – coefficient of variation
MDD – minimum detectable difference
PCB – polychlorinated biphenyl
TTL – target tissue level
ww – wet weight

B5 Clam Tissue

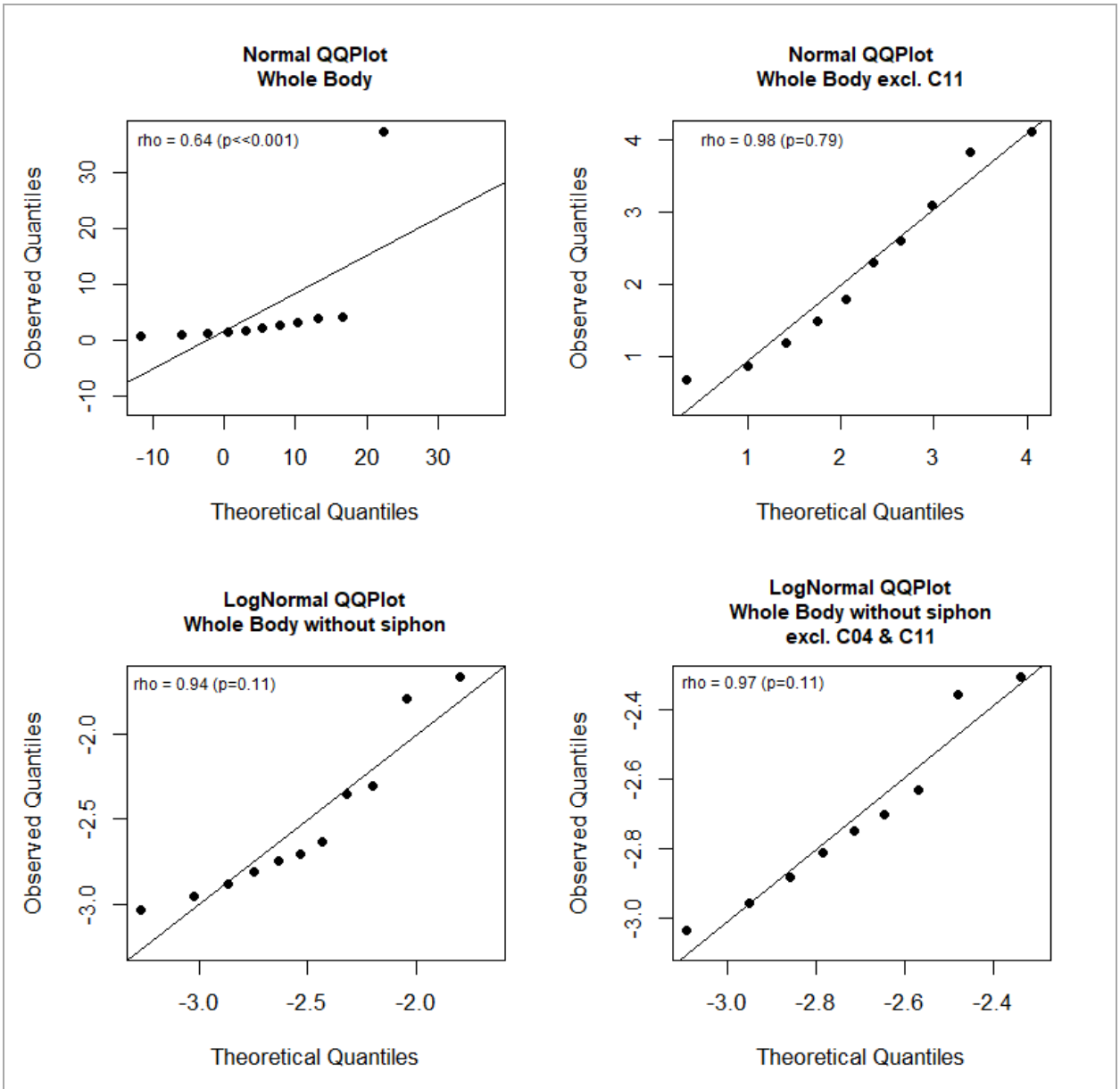
This section provides statistical details regarding the interpretation of the clam tissue data presented in Section 5 of the main report.

B5.1 Influence of Non-Detects

Total PCBs are calculated as the sum of detected Aroclors. For dioxins/furans, the contribution to the total from non-detected compounds ranged from 1 to 60% (using 0.5 method detection limit [MDL]). Individual cPAH compounds below detection contributed 0 to 81% of the total TEQ (using 0.5 MDL). The influence of non-detects on the overall cPAH TEQ led to an analysis of several different treatments of the detection limits in calculation of the TEQ.

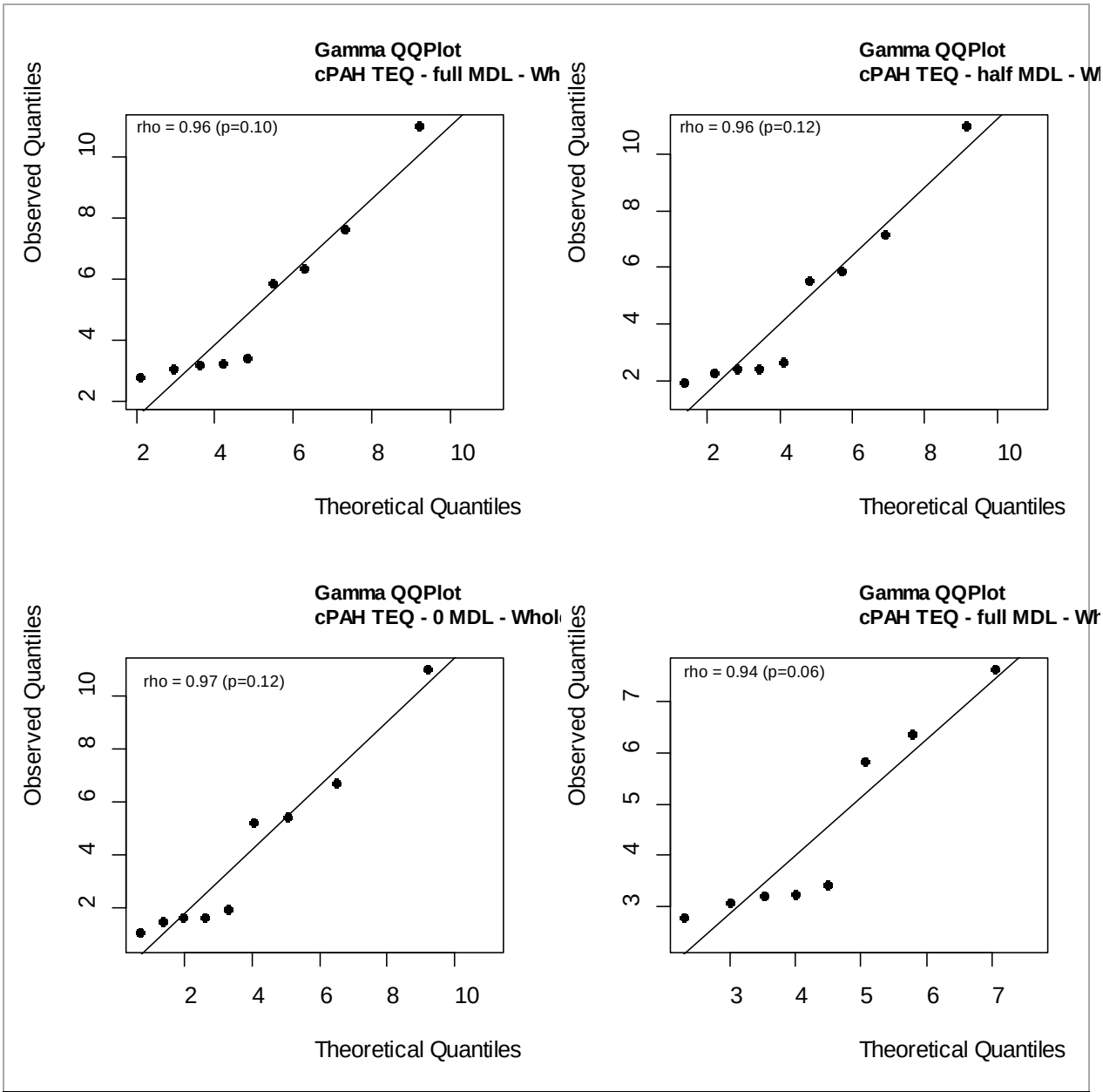
B5.2 95UCL CALCULATIONS

The clam tissue DQO 1 required that the 95UCL for the site-wide mean be established from this dataset for the four risk drivers. Following the methods described in Section 1.1, the best distributional form for each COC was identified in order to use the most appropriate formula for the 95UCL. The best-fitting probability plots are shown in Figures B5-1 through B5-3, and results are summarized in Table B5-1.



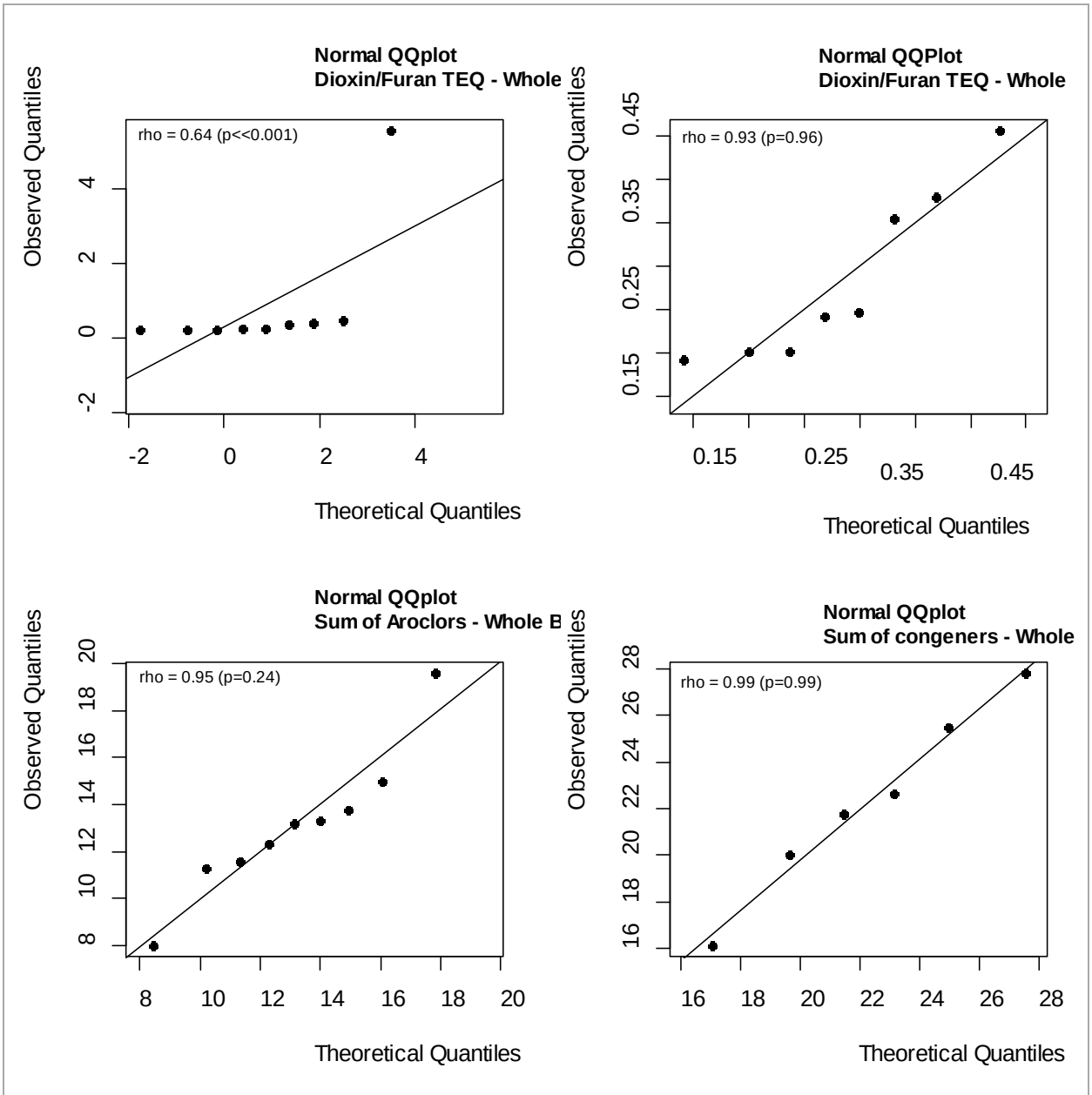
Note: Inorganic arsenic in whole-body (calculated) tissues for all data (top left) and excluding the outlier in area C11 (RM 3.8E) (top right). Inorganic arsenic in whole-body tissues without siphon skin for all data (bottom left) and excluding influential values from areas C04 and C11 (RM 1.5W and RM 3.8E) (bottom left).

Figure B5-1. Probability plots of inorganic arsenic (mg/kg ww) results in clam tissue composite samples



Note: cPAH TEQs in whole-body tissues for all data using full MDL for non-detects (top left), ½ MDL (top right), and zero MDL (bottom left). Results using full MDL and excluding the extreme value in area C05 (Slip 2) (bottom right).

Figure B5-2. Probability plots of cPAH TEQ (µg/kg ww) results in clam tissue composite samples



Note: Dioxin/furan TEQs for all data (top left) and excluding the outlier in area C04 (Glacier Bay) (top right). Total PCB sum of Aroclors (bottom left) and sum of congeners (bottom right).

Figure B5-3. Probability plots for dioxin/furan TEQs (ng/kg ww) and total PCBs (µg/kg ww) in clam tissue composite samples

Table B5-1. Goodness-of-fit and variance statistics for risk drivers in clam tissue composite samples

Risk Driver	n	Best Fit Distribution	PPCC	p-Value	CV
Total PCBs (µg/kg ww)					
Total PCB Aroclors	9	normal	0.95	0.24	0.24
Total PCB congeners	6	normal	0.99	0.99	0.18
Dioxin/furan TEQ (ng/kg ww)					
All data	9	none	0.64	< 0.001	2.0
Excluding outlier from area C04 (Glacier Bay)	8	normal	0.93	0.96	0.35
cPAH TEQ (µg/kg ww)					
Non-detects = MDL	9	gamma	0.96	0.10	0.54
Non-detects = ½ MDL	9	gamma	0.96	0.12	0.67
Non-detects = 0	9	gamma	0.97	0.12	0.84
Excluding highest value (sample from area C05 [Slip 2]); non-detects = MDL	8	gamma	0.94	0.06	0.42
Inorganic arsenic (mg/kg ww)					
Whole body (calculated) (all data)	11	none	0.64	< 0.001	1.98
Whole body (calculated) (excluding outlier from area C11 at RM 3.8E)	10	normal	0.98	0.79	0.54
Whole body without siphon skin	11	lognormal	0.94	0.11	0.54
Whole body without siphon skin – excluding highest values from areas C04 and C11	9	lognormal	0.97	0.11	0.46

cPAH – carcinogenic polycyclic aromatic hydrocarbon
CV – coefficient of variation
MDL – method detection limit
PCB – polychlorinated biphenyl
RM 0 river mile
TEQ – toxic equivalent
ww – wet weight

When individual elevated sample(s) were responsible for the skewness in a distribution, that dataset was also evaluated without the elevated sample(s). The elevated samples were collected from areas expected to be remediated, so the calculations that excluded outliers explored how the data may be expected to behave post-remediation, whereas the baseline 95UCLs are represented by the complete datasets.

For cPAH TEQ, only one of the composite samples had detected concentrations of all cPAH compounds, so treatment of non-detects for the calculation of the TEQ is an important consideration. Three different non-detect assumptions (i.e., equal to the MDL, equal to one-half of the MDL, or equal to zero) were presented in Section 5.2.2.2 of the main report.

The strong influence of non-detects for the cPAHs results in uncertainty about the baseline variance of the cPAH TEQ values in clam tissues. The variance estimates may be different in future monitoring efforts if alternative methods (e.g., high resolution) are

used for cPAH analysis. Because the CVs estimated from the baseline dataset are not expected to be representative of the future variance, the RMEs for future datasets and the MDDs between baseline and future monitoring were not estimated.

B6 References

Apell JN, Gschwend PM. 2017. The atmosphere as a source/sink of polychlorinated biphenyls to/from the Lower Duwamish Waterway Superfund site. Environ Pollut 227:263-270.

DMMP. 2009. *OSV Bold* summer 2008 survey. Data report. The Dredged Material Management Program (DMMP) agencies: US Army Corps of Engineers, Seattle District, Seattle, WA; US Environmental Protection Agency, Region 10, Seattle, WA; Washington State Department of Natural Resources; and Washington State Department of Ecology, Olympia, WA.

Ecology. 2015. Sediment cleanup users manual II. Guidance for implementing the Cleanup Provisions of the Sediment Management Standards, Chapter 173-204 WAC. March 2015. Pub. no. 12-09-057. Toxics Cleanup Program, Washington State Department of Ecology, Olympia, WA.

EPA. 2014. Record of Decision. Lower Duwamish Waterway Superfund Site. US Environmental Protection Agency.

EPA. 2016. Statistical software ProUCL 5.1.00 for environmental applications for data sets with and without nondetect observations [online]. US Environmental Protection Agency, Washington, DC. Updated June 20, 2016. Available from: https://www.epa.gov/land-research/proucl-software.

Millard SP. 2013. EnvStats. An R Package for Environmental Statistics. Springer.

R Core Team. 2018. R: A language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Schuetzenmeister A, Dufey F. 2018. VPF: variance component program [online]. Updated July 18, 2018. Available from: https://cran.r-project.org/web/packages/VCA/index.html.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, NY.

Windward. 2010. Lower Duwamish Waterway remedial investigation. Remedial investigation report. Final. Prepared for Lower Duwamish Waterway Group. Windward Environmental LLC, Seattle, WA.

Windward, Integral. 2017. Pre-design studies work plan. Lower Duwamish Waterway Superfund site. Final. Prepared for the Lower Duwamish Waterway Group for submittal to EPA Region 10 on August 28, 2017. Windward Environmental LLC and Integral Consulting Inc., Seattle, WA.

Windward. 2018a. Lower Duwamish Waterway baseline surface sediment collection and chemical analyses - quality assurance project plan. Final. Windward Environmental LLC, Seattle, WA.

Windward. 2018b. Lower Duwamish Waterway surface sediment data report. Draft final. Submitted to EPA October 9, 2018. Windward Environmental LLC, Seattle, WA.

Windward. [in prep]. Baseline surface water collection and chemical analyses data report. Draft. Submitted to LDWG November 13, 2018. Windward Environmental LLC, Seattle, WA.