APPENDIX D

DATA MANAGEMENT AND INTERPOLATION PARAMETERS

APPENDIX D   DATA MANAGEMENT

D.1   Averaging Laboratory Replicate Samples

Chemical concentrations obtained from the analysis of laboratory replicate samples (two or more analyses of the same sample) will be averaged for a closer representation of the "true" concentration as compared to the result of a single analysis. Averaging rules are dependent on whether the individual results are detected concentrations or reporting limits (RLs) for undetected chemicals. If all concentrations are detected for a single chemical, the values are simply averaged arithmetically for the sample and its associate laboratory replicate sample(s). If all concentrations are undetected for a given parameter, the minimum RL is selected. If the concentrations are a mixture of detected concentrations and RLs, any two or more detected concentrations are averaged arithmetically and RLs ignored. If there is a single detected concentration and one or more RLs, the detected concentration is reported. The latter two rules are applied regardless of whether the RLs are higher or lower than the detected concentration.

D.2   Location Averaging

Results of chemical concentrations of discrete samples collected at a single sampling location that were submitted to the laboratory as individual samples and analyzed separately were averaged for the purposes of mapping a single concentration per location. The averaging rules used for location averaging are the same as for laboratory replicate samples described above. This type of averaging is only performed for mapping purposes. Location averaging is performed in the following instances:

- When multiple sediment samples were collected from the same location at the same time. For example: a sample and its field duplicate, often referred to as a split sample (PSEP 1997).

- When multiple sediment samples were collected at a single location within a short timeframe for monitoring purposes. For example: samples collected at specific sampling locations near the Hanford CSO in 1995 and 1996 were averaged for mapping purposes.

D.3   Significant Figures and Calculations

Analytical laboratories report results with various numbers of significant figures depending on the laboratory's standard operating procedures, the instrument, the chemical, and the reported chemical concentration relative to the RL. The reported (or assessed) precision of each result is explicitly stored in the project database by recording the number of significant figures. Tracking of significant figures is used when calculating analyte sums and performing other data summaries. When a calculation involves addition, such as totaling PCBs, the calculation can only be as precise as the least precise number that went into the calculation. For example:

> 210 + 19 = 229 would be reported as 230 because although 19 is reported to 2 significant digits, the trailing zero in the number 210 is not significant.

When a calculation involves multiplication or division, the final result is rounded at the end of the calculation to reflect the value used in the calculation with the fewest significant figures. For example:

> $59.9 \times 1.2 = 71.88$ would be reported as 72 because there are two significant figures in the number 1.2.

When rounding, if the number following the last significant figure is less than 5, the digit is left unchanged. If the number following the last significant figure is equal to or greater than 5, the digit is increased by 1.

D.4   Best Result Selection for Multiple Results

In some instances, the laboratory generates more than one result for a chemical for a given sample. Multiple results can occur for several reasons, including: 1) the original result did not meet the laboratory's internal quality control (QC) guidelines, and a reanalysis was performed; 2) the original result did not meet other project data quality objectives, such as a sufficiently low RL, and a reanalysis was performed; or 3) two different analytical methods were used for that chemical. In each case, a single best result is selected for use. The procedures for selecting the best result differ depending on whether a single or multiple analytical methods are used for that chemical.

For the same analytical method, if the results are:

- Detected and not qualified, then the result from the lowest dilution is selected, unless multiple results from the same dilution are available, in which case, the result with the highest concentration is selected.

- A combination of estimated and unqualified detected results, then the unqualified result is selected. This situation most commonly occurs when the original result is outside of calibration range, thus requiring a dilution.

- All estimated, then the "best result" is selected using best professional judgment in consideration of the rationale for qualification. For example, a result qualified based on laboratory replicate results outside of QC objectives for precision would be preferred to a qualified result that is outside the calibration range.

- A combination of detected and undetected results, then the detected result is selected. If there is more than one detected result, the applicable rules for multiple results (as discussed above) are followed.

- All undetected results, then the lowest RL is selected.

If the multiple results are from different analytical methods, then the result from the preferred method specified in the quality assurance project plan (QAPP) or based on the consensus of the professional opinions of project chemists was selected.

The following rules are applied to multiple results from different analytical methods:

- For detected concentrations analyzed by the SVOC full-scan and selective ion monitoring (SIM) methods (i.e., PAHs), the highest detected concentration is selected. If the result by one method is detected and the result by the other method is not detected, then the detected result is selected for reporting, regardless of the method. If results are reported as non-detected by both methods, the undetected result with the lowest RL is selected. The SIM method is more analytically sensitive than the full-scan SVOC method, and the undetected results are generally reported at a lower RL by the SIM method than by the full-scan method. Therefore, the SIM method is selected for non-detected results unless an analytical dilution or analytical interferences elevated the SIM RL above the SVOC full-scan RL.

- Hexachlorobenzene and hexachlorocyclopentadiene are analyzed by EPA Methods 8081A, 8270, and/or 8270-SIM. The result from the method with the greatest sensitivity (i.e., lowest RL) is selected if all results are undetected. EPA Method 8081A results are generally selected, when available, because the standard laboratory RLs from this analysis are significantly lower than those from EPA Methods 8270 and 8270-SIM. When chemicals are detected, the detected result with the highest concentration is selected unless the detected concentration is qualified as estimated or tentatively identified, in which case the rule designating treatment of qualified and unqualified data would apply.

D.5   Calculated Totals

Total PCBs, total dichloro-diphenyl-trichloroethane (DDTs), total PAHs, and total chlordane are calculated by summing the detected values for the individual components available for each sample. For individual samples in which none of the individual components is detected, the total value is given a value equal to the highest RL of an individual component, and assigned the same qualifier (U or UJ), indicating an undetected result. Concentrations for the analyte sums are calculated as follows:

- Total PCBs are calculated, in accordance with the methods of the SMS, using only detected values for seven Aroclor mixtures.[1] For individual samples in which none of the seven Aroclor mixtures is detected, total PCBs are given a value equal to the highest RL of the seven Aroclors and assigned a U-qualifier indicating the lack of detected concentrations.

- **Total low-molecular-weight PAHs (LPAHs), high-molecular-weight PAHs (HPAHs), PAHs, and benzofluoranthenes** are also calculated in accordance with the methods of the SMS. Total LPAHs are the sum of detected concentrations for naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, and anthracene. Total HPAHs are the sum of detected concentrations for fluoranthene, pyrene, benzo(a)anthracene, chrysene, total benzofluoranthenes, benzo(a)pyrene, indeno(1,2,3,-c,d)pyrene, dibenzo(a,h)anthracene, and benzo(g,h,i)perylene. Total benzofluoranthenes are the sum of the b (i.e., benzo(b)fluoranthene), j, and k isomers.

---

[1] Aroclors 1016, 1221, 1232, 1242, 1248, 1254, and 1260.

Because the j isomer is rarely quantified, this sum is typically calculated with only the b and k isomers. For samples in which all individual compounds within any of the three groups described above are undetected, the single highest RL for that sample represents the sum.

- **Total DDTs** are calculated using only detected values for the DDT isomers: 2,4'-DDD; 4,4'-DDD; 2,4'-DDE; 4,4'-DDE; 2,4'-DDT; and 4,4'-DDT. For individual samples in which none of the isomers are detected, total DDTs are given a value equal to the highest RL of the six isomers and assigned a U-qualifier, indicating the lack of detected concentrations.

- **Total chlordane** is calculated using only detected values for the following compounds: alpha-chlordane, gamma-chlordane, oxychlordane, cis-nonachlor, and trans-nonachlor. For individual samples in which none of these compounds is detected, total chlordane is given a value equal to the highest RL of the five compounds listed above and assigned a U-qualifier, indicating the lack of detected concentrations.

D.6   Calculation of Carcinogenic Polycyclic Aromatic Hydrocarbons

Carcinogenic polycyclic aromatic hydrocarbons (cPAH) values are calculated using potency equivalency factor (PEF) values (California EPA 2002; Ecology 2001) based on the individual PAH component's relative toxicity to benzo(a)pyrene. PEF values are presented in Table D-1. The cPAH is calculated as the sum of each individual PAH concentration multiplied by the corresponding PEF value. When the individual PAH component concentration is reported as non-detected, then the PEF is multiplied by half the RL.

Table D-1
cPAH PEF Values

cPAH	PEF Value (unitless)
Benzo(a)pyrene	1
Benzo(a)anthracene	0.1
Benzo(b)fluoranthene	0.1
Benzo(k)fluoranthene	0.1
Dibenz(a,h)anthracene	0.4
Indeno(1,2,3-cd)pyrene	0.1
Chrysene	0.01

cPAH – carcinogenic polycyclic aromatic hydrocarbon
PEF – potency equivalency factor

D.7   Calculation of PCB Congener TEQs

PCB congener toxic equivalents (TEQs) are calculated using the World Health Organization (WHO) consensus toxic equivalency factor (TEF) values (Van den Berg et al. 2006) for mammals as presented in Table D-2. The TEQ is calculated as the sum of each congener concentration multiplied by the corresponding TEF value. When the congener concentration is reported as non-detected, then the TEF is multiplied by zero, half the RL or the full RL, depending on the calculation method specified.

Table D-2.
PCB Congener TEF Values for Mammals

PCB Congener Number	TEF Value (unitless)
77	0.0001
81	0.0003
105	0.00003
114	0.00003
118	0.00003
123	0.00003
126	0.1
156	0.00003
157	0.00003
167	0.00003
169	0.03
189	0.00003

PCB – polychlorinated biphenyl
TEF – toxic equivalency factor

D.8   Calculation of Dioxin/Furan Congener TEQs

Dioxin/furan congener TEQs are calculated using the WHO consensus TEF values (Van den Berg et al. 2006) for mammals as presented in Table D-3. The TEQ is calculated as the sum of each congener concentration multiplied by the corresponding TEF value. When the congener concentration is reported as undetected, then the TEF is multiplied by zero, half the RL, or the full RL, depending on the calculation method specified.

Table D-3

Dioxin/Furan Congener TEF Values for Mammals

Dioxin/Furan Congener	TEF Value (unitless)
1,2,3,4,6,7,8-Heptachlorodibenzofuran	0.01
1,2,3,4,6,7,8-Heptachlorodibenzo-*p*-dioxin	0.01
1,2,3,4,7,8,9-Heptachlorodibenzofuran	0.01
1,2,3,4,7,8-Hexachlorodibenzofuran	0.1
1,2,3,4,7,8-Hexachlorodibenzo-*p*-dioxin	0.1
1,2,3,6,7,8-Hexachlorodibenzofuran	0.1
1,2,3,6,7,8-Hexachlorodibenzo-*p*-dioxin	0.1
1,2,3,7,8,9-Hexachlorodibenzofuran	0.1
1,2,3,7,8,9-Hexachlorodibenzo-*p*-dioxin	0.1
1,2,3,7,8-Pentachlorodibenzofuran	0.03
1,2,3,7,8-Pentachlorodibenzo-*p*-dioxin	1
2,3,4,6,7,8-Hexachlorodibenzofuran	0.1
2,3,4,7,8-Pentachlorodibenzofuran	0.3
2,3,7,8-Tetrachlorodibenzofuran	0.1
2,3,7,8-Tetrachlorodibenzo-*p*-dioxin	1
Octachlorodibenzofuran	0.0003
Octachlorodibenzo-*p*-dioxin	0.0003

TEF – toxic equivalency factor

D.9 TOC Normalization

Sediment samples with TOC contents < 0.5% or > 4% are not TOC normalized for comparison to the organic carbon normalized SMS criteria. The lower bound value of 0.5% TOC is provided by Ecology (Michelsen and Bragdon-Cook 1993). The upper bound value of 4% was consistent with the value used in the LDW RI. When TOC normalization was not possible and the dry-weight concentration was > LAET and ≤ 2LAET, the concentration was considered to be > SQS and ≤ CSL.

D.10 References

California EPA. 2002. Air Toxics Hot Spots Program Risk Assessment Guidelines, Part II Technical Support Document for Describing Available Cancer Potency Factors. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency.

Ecology. 2001. Model Toxics Control Act Cleanup Regulation, Chapter 173-340 WAC. Publication No. 94-06. Toxics Cleanup Program, Washington State Department of Ecology, Olympia, WA.

Michelsen TC, Bragdon-Cook K. 1993. Technical information memorandum: Organic carbon normalization of sediment data. Washington Department of Ecology, Olympia, WA.

PSEP. 1997. Recommended guidelines for sampling marine sediment, water column, and tissue in Puget Sound. Prepared for the Puget Sound Estuary Program, US Environmental Protection Agency, Region 10. King County (METRO) Environmental Laboratory, Seattle, WA.

Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, Fiedler H, Hakansson H, Hanberg A, Haws L, Rose M, Safe S, Schrenk D, Tohyama C, Tritscher A, Tuomisto J, Tysklind M, Walker N, Peterson RE. 2006. The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicol Sci 93(2):223-241.

APPENDIX D-2 EW INTERPOLATION PARAMETERS

WindWard environmental LLC

MEMORANDUM

From: Susan McGroddy and Craig Hansen Windward Environmental, on behalf of the Port of Seattle

Subject: EW SRI IDW interpolations

Date: May 21, 2012

Cc: Doug Hotchkiss, Jeff Stern, Debra Williston, Pete Rude, Tom Wang, Dan Berlin

---

The interpolation of the surface sediment contaminant concentrations for East Waterway (EW) presented in the EW SRI was conducted using the interpolation methods developed for the Lower Duwamish Waterway (LDW) as was specified in the EW SRI Work Plan (Anchor and Windward 2007). The interpolation was an inverse distance weighting (IDW) method. The specific input parameters are provided in Table 1.

**Table 1. Input parameters for EW and LDW IDW interpolations**

Parameter	EW	LDW
Search radius	150 ft with a minimum of one neighbor	150 ft with a minimum of one neighbor
Power	5	5
Cell size	10ft	10ft
Input barrier	yes	no

The input parameters for the IDW interpolations were the same for EW and LDW with the exception of the input barrier which was used in EW and not in LDW. The MHHW polyline was used as an input barrier for EW. The MHHW study area boundary polygon was used as a mask to restrict the valid values of the interpolation to the in-water portion of the study area. The input barrier stops the interpolation at the shoreline which prevents values separated by dry land from influencing one another. The application of an input barrier increases the reliability of the interpolation.

For each contaminant, the surface sediment database was queried by chemical name to create the subset of relevant locations, which were plotted in ESRI ArcGIS 10 by their X

and Y coordinates. The raster surface was interpolated from the Value_or_HalfQL value at these locations using the IDW tool in Spatial Analyst.

An Average Nearest Neighbor analysis of surface sediment locations reveals an evenly dispersed sample pattern in the EW compared to a random dispersal pattern in the LDW (Figure 1). The spatial distribution suggests the dataset is well-suited to IDW interpolation. The IDW with the specified input parameters should be as reliable or more reliable for EW compared to LDW.
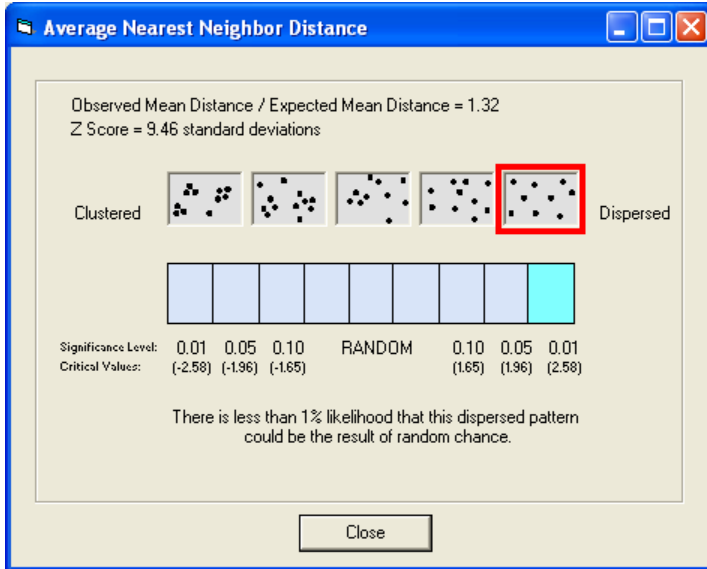
EW surface sediment locations
```
Average Nearest Neighbor
Summary
Observed Mean Distance:
106.788796
Expected Mean Distance:
80.738260
Nearest Neighbor Ratio:
1.322654
Z Score:        9.462417
p-value:        0.000000
Minimum 1 neighbor distance:
0.000000
Average 1 neighbor distance:
106.788796
Maximum 1 neighbor distance:
297.701064
```
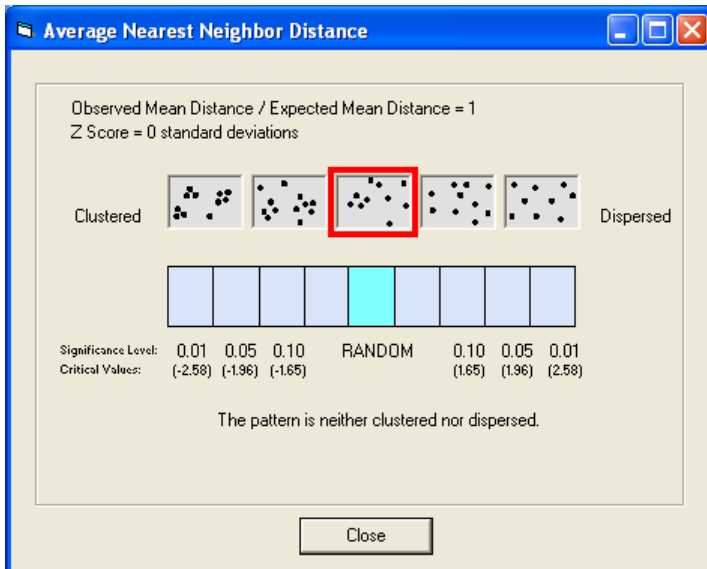

LDW surface sediment locations
```
Average Nearest Neighbor
Summary
Observed Mean Distance:
60.594528
Expected Mean Distance:
60.597285
Nearest Neighbor Ratio:
0.999954
Z Score:        -0.003217
p-value:        0.997433
Minimum 1 neighbor distance:
1.000000
Average 1 neighbor distance:
60.594528
```

Figure 1. Nearest Neighbor Analysis results for EW and LDW

References

LDWG (Lower Duwamish Waterway Group), 2007. Draft Memorandum: Updated
    Methodology for Interpolating Surface Sediment Chemistry in the Lower
    Duwamish Waterway Feasibility Study. Prepared by RETEC. December 11.

Additional IDW Parameterization

Following the completion of the draft EW SRI, an additional analysis was conducted to identify optimized IDW parameters based on the EW sediment dataset. Specifically, point data used to create IDW surfaces included samples from the entire study area with PCB results in surface sediment (0 to 10 centimeters [cm]) and shallow subsurface sediment (0 to 2 feet), which included the maximum result for sediment core samples in the upper 2 feet below mudline north of the Spokane Street Bridge. Also included in the query were 0- to 10-cm samples collected following dredging in 2005 in the Phase 1 removal area prior to placement of clean cover material.

ESRI's ArcGIS Geostatistical Analyst (GA) was used to create the IDW surfaces using the input parameters for circular search radius and power fixed to the values shown in Table 1. Consistent with the method used in the LDW Memo, the maximum/minimum number of closest samples used for grid-cell interpolation was varied between 1/1 and 10/1. Cell size was set at 10 feet, and mean higher high water (MHHW) was used as an input barrier to prevent interpolation between areas separated by dry land.

In order to evaluate the errors of each parameter set, both a GA layer and an ESRI grid were created. The cross-validation tool available within GA was used to calculate the mean error and the root mean square error (RMSE).

The mean error can be defined as the averaged difference between the measured and predicted values and calculated by the equation below.

$$\frac{\sum_{i=1}^{n}\left(\widehat{Z}(s_i) - z(s_i)\right)}{n}$$

where:

n	=	number of points
$\widehat{Z}$	=	measured value
z	=	predicted value
s	=	value
i	=	point number

The RMSE is the square root of the averaged squared difference between the measured and predicted values and determined by the equation below.

$$\sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{Z}(s_i) - z(s_i)\right)^2}{n}}$$

where:

n	=	number of points
$\hat{Z}$	=	measured value
z	=	predicted value
s	=	value
i	=	point number

Cross-validation calculates error by omitting a point from the input, calculating the interpolated value using the remaining points, and then comparing the interpolated value to the measured value. This is conducted for each point in the dataset to determine the mean error and RMSE. In addition, a point table was exported for each IDW from GA, which included the measured and interpolated value for each point, and was subsequently used to calculate the mean absolute error.

In addition to the cross-validation errors, an observed RMSE was also calculated. The observed RMSE was calculated in the same manner as in the LDW Memo and was used along with the RMSE to identify the optimized set of interpolation parameters. Observed RMSE is calculated using the same RMSE equation; however, points are not iteratively removed. Rather, the difference between the measured and predicted values at each point location is used. Results may differ from the CV RMSE if individual data points are not spatially coincident with the IDW raster cells, which is a function of the point distribution and raster cell size and extent. To facilitate the calculation of observed RMSE, a simple process was built within ArcGIS Model Builder to automate the geoprocessing.

Consistent with the process described in the LDW Memo, the lowest RMSE and observed RMSE were the key statistical metrics used to identify the optimized set of parameters for IDW interpolation in the EW. The parameter combination with the lowest RMSE has the lowest dataset variability. RMSE decreases as the search radius increases and as the power decreases (within each search radius group). The IDW interpolation with the lowest observed RMSE results in the lowest error based on a comparison of measured versus predicted values. Based on these metrics, parameters for IDW interpolation using the EW FS dataset are optimized with a power of 1 and circular search radius of 75 feet, as indicated in Table 1.

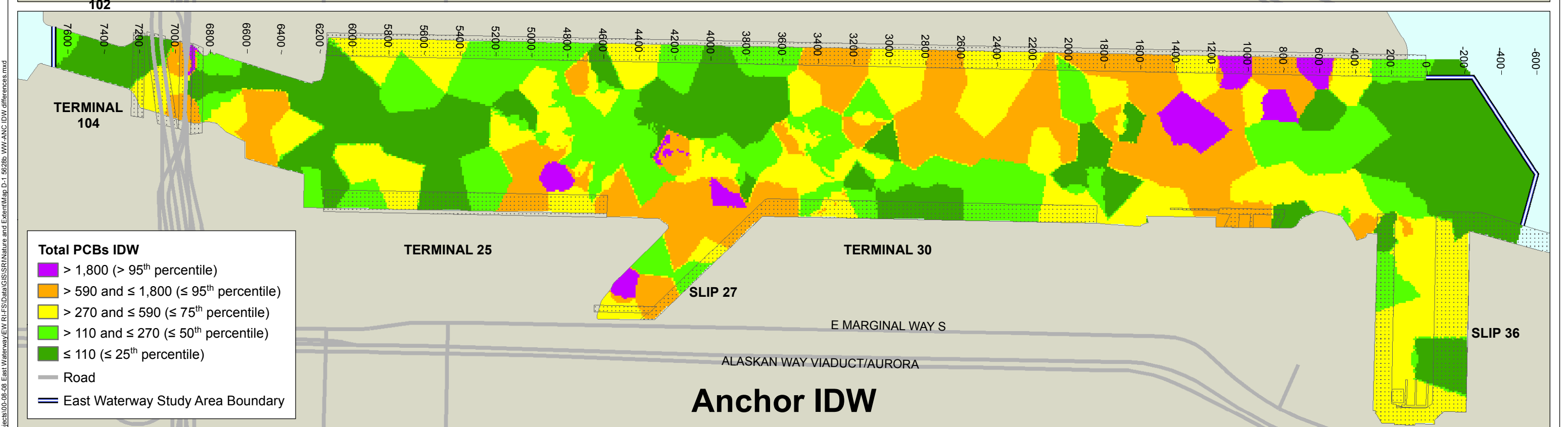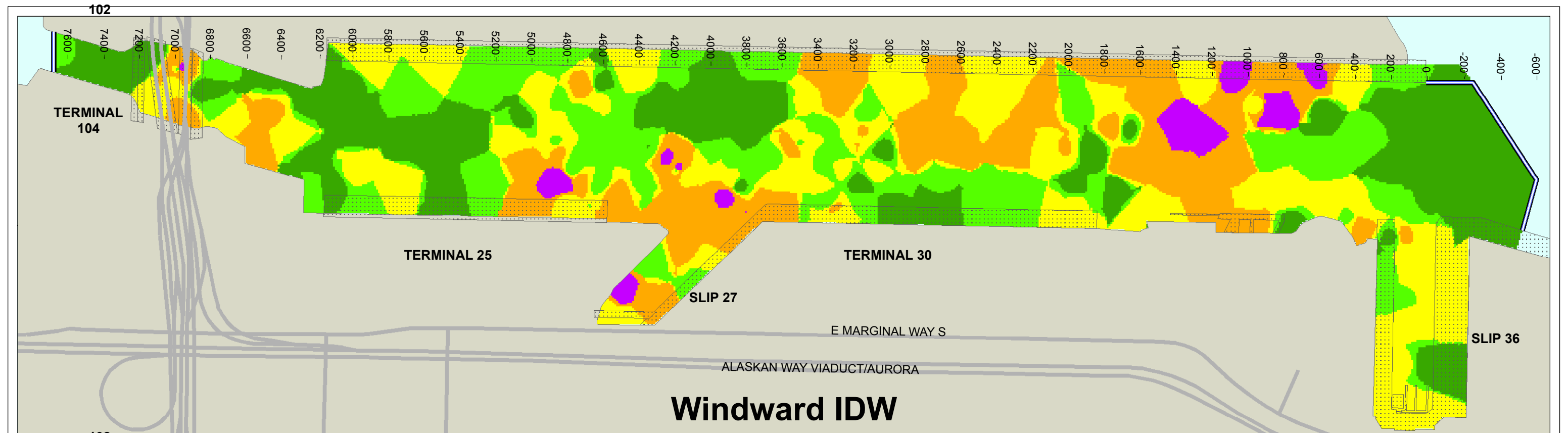**Table 1   Interpolation Parameters Tested for Total PCBs – East Waterway**

| Power | Circular Search Radius (feet) | Cross Validation | | | Observed RMSE |
		Mean Error	Mean Absolute Error	RMSE	
1	250	57.3	714.4	1,260	675
2	250	76.3	750.7	1,351	666
3	250	80.2	766.6	1,406	728
4	250	80.8	776.2	1,438	793
5	250	81.2	784.5	1,456	851
10	250	86.6	809.8	1,506	1,000
1	150	99.0	811.9	1,432	670
2	150	106.6	826.9	1,487	666
3	150	105.8	829.3	1,519	728
4	150	103.6	830.0	1,536	793
5	150	101.8	832.0	1,547	851
10	150	101.7	841.5	1,578	1,000
1	75	94.9	878.0	1,625	648
2	75	95.7	878.3	1,636	666
3	75	95.3	878.8	1,638	728
4	75	95.2	872.8	1,640	793
5	75	95.6	873.0	1,642	851
10	75	100.5	876.6	1,656	1,000

Notes:

1. A maximum of 10 and a minimum of 1 "nearest neighbor" data points were used in all interpolations.

2. Cell size for all interpolations is 10 feet.

3. Lowest Observed RMSE occurs with power of 1 and circular search radius of 75 feet (shaded).

PCB – polychlorinated biphenyl

RMSE – root mean square error

The IDW interpolation for PCBs was run using the parameters used in the SRI which is labeled Windward IDW and the parameters that were selected based on the optimization discussed above which is labeled Anchor IDW. The two interpolations are provided in Map D-1. The interpolations are very similar which is consistent with the spatially well-distributed surface sediment dataset.

Windward IDW

- 102
- 7600, 7400, 7200, 7000, 6800, 6600, 6400, 6200, 6000, 5800, 5600, 5400, 5200, 5000, 4800, 4600, 4400, 4200, 4000, 3800, 3600, 3400, 3200, 3000, 2800, 2600, 2400, 2200, 2000, 1800, 1600, 1400, 1200, 1000, 800, 600, 400, 200, 0, -200, -400, -600

TERMINAL 104

TERMINAL 25

TERMINAL 30

SLIP 27

SLIP 36

E MARGINAL WAY S

ALASKAN WAY VIADUCT/AURORA

Anchor IDW

- 102
- 7600, 7400, 7200, 7000, 6800, 6600, 6400, 6200, 6000, 5800, 5600, 5400, 5200, 5000, 4800, 4600, 4400, 4200, 4000, 3800, 3600, 3400, 3200, 3000, 2800, 2600, 2400, 2200, 2000, 1800, 1600, 1400, 1200, 1000, 800, 600, 400, 200, 0, -200, -400, -600

TERMINAL 104

TERMINAL 25

TERMINAL 30

SLIP 27

SLIP 36

E MARGINAL WAY S

ALASKAN WAY VIADUCT/AURORA

Total PCBs IDW

- > 1,800 (> 95th percentile)
- > 590 and ≤ 1,800 (≤ 95th percentile)
- > 270 and ≤ 590 (≤ 75th percentile)
- > 110 and ≤ 270 (≤ 50th percentile)
- ≤ 110 (≤ 25th percentile)
- Road
- East Waterway Study Area Boundary

WindWard environmental LLC

N

| 0 | 300 | 600 |
Feet

| 0 | 90 | 180 |
Meters
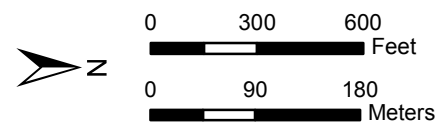
Map D-1. Difference between WW IDW and Anchor IDW for total PCBs