Development of Benthic SQVs for Freshwater Sediments in Washington, Oregon, and Idaho

June, 2010

Avocet Consulting

Development of Benthic SQVs for Freshwater Sediments in Washington, Oregon, and Idaho

June, 2010

Prepared for:

Washington Department of Ecology
Olympia, WA

and

Oregon Department of Environmental Quality
Portland, OR

Prepared by:
Teresa Michelsen, Ph.D.
Avocet Consulting
Olympia, WA

Under contract to:
Ecology & Environment
Seattle, WA

ACKNOWLEDGMENTS

This report was prepared by Dr. Teresa Michelsen, under contract to E&E and the Washington Department of Ecology, on behalf of the RSET agencies. Dr. Michelsen's work was guided by the RSET Sediment Quality Guidelines Workgroup, whose members were as follows:

Keith Johnson, Oregon DEQ, Chair
Mike Anderson, Oregon DEQ (retired)
Robert Anderson, NOAA
Jeremy Buck, USF&W
Taku Fuji, Kennedy Jenks
Dan Gambetta, NOAA
Laura Inouye, Washington Dept. of Ecology
Lyndal Johnson, NOAA
Mike Poulsen, Oregon DEQ
Paul Seidel, Oregon DEQ
Burt Shephard, EPA
Mark Siipola, Portland COE
Dave Sternberg, Washington Dept. of Ecology

Invaluable contract assistance was provided by Bill Richards, Blythe Mackey, and Erin Lynch of Ecology & Environment, and James J. McAteer Jr. of QA/QC Solutions. Mike Anderson performed the coding of the FPM method in addition to participating in the workgroup.

Also instrumental in guiding and supporting this process was the RSET management team and sediment staff at Ecology, including Jim Reese (NW Region Corps), Stephanie Stirling (Seattle District Corps), Dave Bradley, Chance Asher, and Russ McMillan (WA Dept. of Ecology), and Cathy Tortorici (NOAA). Funding was provided by the Oregon Department of Environmental Quality, Washington Department of Ecology, and NW Region Corps of Engineers.

TABLE OF CONTENTS **Page**

Executive Summary **ES-1**

1. Introduction **1**

1.1 Freshwater SQV Early Development 1

1.2 2007 Update of the Freshwater SQVs 2

1.3 Public Outreach and Peer Review 2

1.4 Report Organization and Purpose 3

2. Methods **4**

2.1 Database Development 4

2.1.1 Data Collection 4

2.1.2 Initial Data Screening 4

2.1.3 Normalization and Summing 6

2.1.4 Comparison to Control vs. Reference 7

2.1.5 Bioassay Tests and Endpoints 7

2.1.6 ANOVA Analyte Screening 11

2.2 SQV Calculation – Floating Percentile Method 11

2.3 Reliability Analysis 15

2.4 Exploratory Model Runs 16

2.5 Final Model Runs 17

2.6 Decision Framework 18

2.6.1 Regulatory Considerations 18

2.6.2 Technical Approach 19

3. Results **20**

3.1 Final Data Set 20

3.2 Reliability Assessment 20

3.3 Sediment Quality Guidelines 27

4. Conclusions **33**

5. References **34**

Appendix A. List of Surveys.....A-1

Appendix B. Data Screening.....B-1

TABLES **Page**

Table ES-1. Recommended Sediment Quality Guidelines..... ES-3
Table 2-1. Qualifier Definitions for Screened-out Data..... 5
Table 2-2. Quality Assurance and Adverse Effects Levels for Biological Tests..... 8
Table 2-3. Reliability Goals for Proposed Freshwater SQVs..... 16
Table 3-1. Bioassays and Endpoints in the Final Data Set..... 20
Table 3-2. Biological Hits at Each Effects Level..... 20
Table 3-3. Reliability of the FPM Results and Existing SQV Sets at the SQS/SL1 Level..... 22
Table 3-4. Reliability of the FPM Results and Existing SQV Sets at the CSL/SL2 Level..... 25
Table 3-5. Floating Percentile Model Values at the SQS/SL1 Level..... 29
Table 3-6. Floating Percentile Model Values at the CSL/SL2 Level..... 30
Table 3-7. Selection of Recommended Sediment Quality Guidelines..... 31
Table A-1. List of Surveys..... A-1
Table B-1. Rarely Detected Analytes..... B-3
Table B-2. ANOVA Screening..... B-7
Table B-3. Summary of Screened Analytes..... B-9

FIGURES **Page**

Figure 2-1. Floating Percentile Method.....13
Figure 2-2. Reliability Measures..... 15

LIST OF ACRONYMS

AETs – Apparent Effects Thresholds
ANOVA – Analysis of variance
ASTM – American Society for Testing and Materials
CSL – Cleanup Screening Level
DDD/DDE/DDT – dichlorodiphenyldichloroethane/dichlorodiphenyldichloroethylene/ dichlorodiphenyltrichloroethane
DEQ – Oregon Department of Environmental Quality
DMEF – Dredged Material Evaluation Framework
DMMP – Dredged Material Management Program
Ecology – Washington Department of Ecology
EIM – Environmental Information Management System
EPA – United States Environmental Protection Agency
ERL – Effects Range – Low
ERM – Effects Range - Median
ESA – Endangered Species Act
FPM – Floating Percentile Model
ID – State of Idaho
LEL – Low Effects Level
MTCA – Model Toxics Control Act
NOAA – National Oceanic and Atmospheric Administration
OR – State of Oregon
PAHs – Polynuclear aromatic hydrocarbons
PCBs – Polychlorinated biphenyls
PEC – Probable Effects Concentration
PEL – Probable Effects Level
PSEP – Puget Sound Estuary Program
QA/QC – Quality assurance/quality control
QA2 – Quality assurance level 2 (litigation/regulation quality)
RSET – Regional Sediment Evaluation Team
SEDQUAL – Sediment Quality database
SEF – Sediment Evaluation Framework
SEL – Severe Effects Level
SETAC – Society for Toxicology and Chemistry
SL1/SL2 – Screening Level 1 or 2
SMARM – Sediment Management Annual Review Meeting
SMS – Sediment Management Standards
SQVs – Sediment quality guidelines
SQS – Sediment Quality Standard
TEC – Threshold Effects Concentration
TEL – Threshold Effects Level
TEQ – Toxicity equivalency quotient
TPH – Total petroleum hydrocarbons
USF&W – United States Fish and Wildlife Service
WA – Washington State

EXECUTIVE SUMMARY

In early 2002, the Washington State Department of Ecology (Ecology) embarked on a project to identify, update, and ultimately select freshwater sediment quality guidelines (SQVs) for use in Ecology's sediment management programs. This effort was completed in July 2003 (SAIC and Avocet 2003), and included compilation of freshwater sediment data in western Washington and Oregon, identification of existing freshwater SQVs in North America, an assessment of their reliability in predicting effects in Washington State, and calculation of alternative SQVs with greater reliability using the Floating Percentile Model (FPM).

The 2003 Ecology database allowed calculation of four acute and subchronic SQVs (*Hyalella* 10-day mortality, *Chironomus* 10-day mortality, *Chironomus* 10-day growth, and Microtox®) using the FPM. There were not enough data for benthic community indices or chronic freshwater tests to enable calculation of chronic SQVs at that time. There was also a lack of data for areas east of the Cascades, and for a variety of pesticides, herbicides and biocides, among other chemicals.

In 2007, the Regional Sediment Evaluation Team (RSET) decided to update Ecology's freshwater SQVs for inclusion in the Sediment Evaluation Framework (SEF) for Oregon, Washington, and Idaho. The SEF is used to evaluate dredging and cleanup projects in marine waters and freshwater areas of these three states, and RSET includes a wide variety of federal and state agencies responsible for these regulatory functions. In 2009, RSET endorsed revisions to the SEF that included interim Freshwater SQVs based on this work. In addition, in 2009, Ecology supported completion of this report as part of the update of the Sediment Management Standards (SMS) and Model Toxics Control Act (MTCA).

The primary goals of the update described in this report were to:

- Include data from a broader geographic area, including areas east of the Cascades and all three states
- Include a broader range of chemicals
- Include at least two chronic tests
- Include several large data sets from recent state and federal cleanup projects, as well as many smaller recent data sets from dredging and cleanup projects
- Obtain consensus among the RSET agencies on how the SQV calculations and reliability analysis should be conducted, along with the final values
- Automate the FPM process so that any of the agencies or stakeholders could make use of it and update the SQVs in the future

The freshwater data set is considerably larger and more diverse in terms of both chemistry and bioassays than it was in 2003, and has been improved from a quality assurance standpoint. The current database allows calculation of FPM values for five acute and chronic endpoints. All data included in the data set were collected using ASTM- and Ecology-approved bioassay methods and chemistry analytical techniques. The data have been validated to a level suitable for regulation and litigation, known as QA2.

The data were collected from western Washington and Oregon and from eastern Washington. No data were identified in eastern Oregon or Idaho that included both bioassay and chemistry data. The data set encompasses a wide variety of different types of environments, including large and small lakes on both sides of the Cascades, large rivers on both sides of the Cascades such as the Duwamish, Willamette,

Columbia, and Spokane Rivers, and small streams. Each data set represents field-collected samples with both chemistry and bioassay data collected at the same time and place.

The following conclusions can be drawn based on the work presented in this report:

- **Accuracy.** Use of the floating percentile method resulted in SQVs that were able to accurately identify 75-80% of the toxic samples, 65-95% of the non-toxic samples, and overall, correctly predicted bioassay results 70-85% of the time (depending on the specific test and endpoint).

- **Comparison to Existing SQVs.** The FPM values represent a substantial improvement in accuracy in identifying non-toxic samples than other available SQV sets, greatly improving the efficiency and implementability of the SQVs. In addition, at the higher effects levels, the FPM values are more able to detect toxic samples than other existing SQV sets.

Based on the conclusions above and an approach developed by the interagency workgroup for combining the individual endpoint values, SQVs for both the SQS/SL1 and the CSL/SL2 levels are recommended (Table ES-1). The method used to develop these values is based on specific assumptions about the levels of risk and error that can be tolerated at each effects level, and provides the opportunity for revision of the SQVs if alternative policy choices are made during the agency and public review process.

It should be reiterated that these values were developed to protect only against toxicity to the benthic community. They are not protective of bioaccumulative effects to humans, wildlife, or fish. However, based on a review and consultation with NOAA and USF&W regarding endangered species in WA, OR, and ID, they are expected to be protective of endangered benthic species.

Additional information for site managers is included in Appendix B, including a list of chemicals that were screened out and the reasons for doing so, and how to evaluate chemicals that do not have recommended SQVs.

Table ES-1. Recommended Sediment Quality Guidelines

Analyte	SQS/SL1[a]	CSL/SL2[b]
Conventional Pollutants (mg/kg)		
Ammonia	230	300
Total sulfides	39	61
Metals (mg/kg)		
Antimony	0.3	12
Arsenic	14	120
Cadmium	2.1	5.4
Chromium	72	82
Copper	400	1200
Lead	360	> 1300
Mercury	0.66	0.8
Nickel	26	110
Selenium	11	> 20
Silver	0.58	1.7
Zinc	3200	> 4200
Organic Chemicals (µg/kg)		
4-Methylphenol	260	2000
Benzoic acid	2900	3800
beta-Hexachlorocyclohexane	7.2	11
bis(2-Ethylhexyl)phthalate	500	22000
Carbazole	1100	1400
Dibenzofuran	200	680
Dibutyltin	910	130000
Dieldrin	4.9	9.3
Di-n-butyl phthalate	380	450
Di-n-octyl phthalate	39	> 1100
Endrin ketone	8.5	**
Monobutyltin	540	> 4800
Pentachlorophenol	1200	> 1200
Phenol	120	210
Tetrabutyltin	97	> 97
Total Aroclors	110	2500
Total DDDs	310	860
Total DDEs	21	900
Total DDTs	100	8100
Total PAHs	17000	30000
Tributyltin	48	320
Bulk Petroleum Hydrocarbons (mg/kg)		
TPH-Diesel	340	510
TPH-Residual	3600	8400

[a] Sediment Quality Standard/Screening Level 1
[b] Cleanup Screening Level/Screening Level 2
> "greater than" value indicates that the toxic level is unknown, but above the concentration shown. If concentrations above this level are encountered, bioassays should be run to evaluate the potential for toxicity.
** No SQV could be set due to limited data above the SL1 concentration.

1.0 INTRODUCTION

This report presents the results of the 2010 recalculation of freshwater sediment quality guidelines (SQVs) for Washington, Oregon, and Idaho. The SQVs were updated by a Regional Sediment Evaluation Team (RSET) workgroup for inclusion in the Sediment Evaluation Framework (SEF) for Oregon, Washington, and Idaho. The SEF is used to evaluate dredging and cleanup projects in both marine waters and freshwater areas of these three states, and RSET includes a wide variety of federal and state agencies responsible for these regulatory functions. In addition, the Washington Department of Ecology supported development of these SQVs for use in cleaning up contaminated sediment sites under the Sediment Management Standards (SMS).

1.1 Freshwater SQV Early Development

In early 2002, Ecology embarked on a project to identify, update, and recalculate freshwater SQVs for use in Washington State sediment management programs. Two levels of SQVs were developed, corresponding to the SMS narrative Sediment Quality Standard (SQS) and Cleanup Screening Level/Minimum Cleanup Level (CSL/MCUL). In the RSET dredging programs, these levels are referred to as Screening Levels 1 and 2 (SL1 and SL2), respectively. Both designations will be used in this report.

Phase I of the project was completed in December 2002 (SAIC and Avocet 2002), and included:

- An update of the regional freshwater sediment database, including gathering additional synoptic data sets, and conducting quality assurance reviews of both new and old data sets.
- Adding new freshwater bioassay evaluation tools to Ecology's SEDQUAL sediment database and analytical tool, allowing the development of custom bioassay hit/no-hit definitions and comparison of bioassay data to those definitions to identify stations with hits.
- A reliability analysis of eight existing SQV sets against the newly updated freshwater data set, to evaluate their ability to correctly predict biological hits and no-hits.
- An evaluation of the use of marine Apparent Effects Thresholds (AETs) as freshwater dredged material disposal guidelines, and recommended updates to the Columbia River Dredged Material Evaluation Framework (DMEF 1998).

The results of these 2002 analyses indicated that existing freshwater SQV sets were not able to correctly predict both hits and no-hits with an acceptable degree of reliability, and further work was therefore needed in Phase II to calculate new freshwater SQVs. Phase II, completed in June 2003, included the following activities (SAIC and Avocet 2003):

- Calculation of alternative freshwater SQVs, based on an iterative error rate minimization technique known as the Floating Percentile Model (FPM).
- A reliability analysis of the FPM SQVs based on the updated regional freshwater data set.
- Recommendations for how these values could be used in Ecology's programs.

This effort produced interim values of good reliability that were applicable to western Washington and Oregon. The interim freshwater SQVs were published and used as guidance by Ecology on a site-specific basis, but have not yet been promulgated. While the overall reliability was high (approximately 80%) and error rates were low (< 20% false negatives and false positives), the data set did not have a geographic scope that encompassed the entire state and did not include chronic tests due to lack of sufficient chronic data at the time.

1.2 2007 Update of the Freshwater SQVs

In 2007, RSET undertook an update of Ecology's freshwater SQVs for inclusion in the SEF. The primary goals of the update described in this report were to:

- Include data from a broader geographic area, including areas east of the Cascades and all three states (WA, OR, ID).
- Include a broader range of chemicals.
- Include at least two chronic tests.
- Include several large data sets from recent state and federal cleanup projects, as well as many smaller recent data sets from dredging and cleanup projects.
- Obtain consensus among the RSET agencies on how the SQV calculations and reliability analysis should be conducted, along with the final values.
- Automate the FPM process so that any of the agencies or stakeholders could make use of it and update the SQVs in the future.

To complete these tasks, an SQV Workgroup was formed and met throughout 2007-2008 to guide the development effort. Members of the workgroup are listed in the acknowledgments, and included federal and state agency representatives and contractors. The final values associated with the workgroup process were calculated in 2008. However, the calculations indicated that the results for two of the most widely used acute mortality bioassays did not meet the workgroup's reliability goals, and consensus was not reached on how to proceed with final development of SQVs.

In 2009, Ecology began an update of the Sediment Management Standards (SMS) and the Model Toxics Control Act (MTCA) regulations. As part of this process, Ecology and the Oregon Department of Environmental Quality (DEQ) agreed to recalculate the results for these two bioassays using alternative effects thresholds recommended by agency technical staff, the SMS Workgroup (an external advisory group for the SMS rule revisions), regional laboratories, and national SQV experts. This approach resulted in SQVs with improved reliability and a complete set of acute and chronic endpoints with reliable SQVs. The results of these combined group efforts are included in this report.

1.3 Public Outreach and Peer Review

The RSET program sought regional and national sediment experts to review and critique the development of freshwater SQVs based on this method. In addition to RSET and the SQV Workgroup, the modeling approach used in the FPM and its results have been presented at numerous conferences, workshops, and public meetings to date, including:

- 1999 SETAC North America Conference, Philadelphia, PA
- 2001 Peer review and public demonstrations of the model in Portland and Seattle as part of the Oregon DEQ Portland Harbor site investigation
- 2003 Sediment Management Annual Review Meeting (SMARM), Seattle, WA
- 2004 SETAC North America Conference, Portland, OR
- 2008 Advanced Sediment Cleanup Conference, Seattle, WA
- 2008, 2009, and 2010 RSET/SMARM public meetings in Seattle, Boise, Portland, and Vancouver
- 2009 Battelle International Conference on Remediation of Contaminated Sediments, Jacksonville, FL
- 2009 PNW-SETAC Conference, Port Townsend, WA

Additional formal public review and comment will occur as part of Ecology's SMS advisory group process and public meetings associated with the SMS rule revision. Once this report is approved by Ecology, an associated journal article will be written and submitted to *Integrated Environmental Assessment and Management*, a peer-reviewed SETAC journal.

1.4 Report Organization and Purpose

Section 2 of this report describes the methods used to update and process the data set, calculate the SQVs, and conduct the reliability assessment. Section 3 presents the updated SQVs and the associated reliability analyses. Section 4 summarizes conclusions and recommendations, and Section 5 provides the references for the report.

It should be emphasized that this report provides recommendations to Ecology and the other RSET agencies, who will make the final decision on how any SQVs presented in this report, or any modifications to the SQVs presented here, will actually be used in state and federal sediment management programs. Additionally, the SQVs presented in this report were guided and based on initial policy and technical decisions made by RSET and refined by Ecology and DEQ, discussed in further detail in section 2.6. Any potential future modifications to these underlying choices and conditions could significantly change the associated values. Final selected values will appear in a revised version of the SEF and/or the SMS.

2.0 METHODS

2.1 Database Development

The following sections describe the collection, screening, processing, and assembly of the data set used in the FPM model runs.

2.1.1 Data Collection

The data set for this effort includes most of the data originally collected by Ecology in 2002-2003 (see SAIC and Avocet 2002, 2003 for details), although some of those original data were excluded during this effort because they did not use modern protocols or had fewer replicates than are currently required. Additional data collection was conducted in 2007 to obtain data sets from a broader geographic region (all areas of OR, WA, and ID), data sets with chronic bioassays, and more recent data. Data collection efforts continued for approximately one year, and were largely successful in meeting the project goals, as follows:

- The size of the overall data set was approximately tripled from the 2003 data set.
- Data sets were included from east of the Cascades in Washington State.
- The data set includes many analytes not well represented in the 2003 data set.
- Several recent, large studies of special interest to the agencies were included, including Willamette River, Portland Harbor, Upper Columbia River, and Spokane River studies.
- Substantial chronic data was obtained for the *Hyalella azteca* 28-day growth and mortality endpoints.

Several goals of the data collection effort were not able to be met. No studies with complete analyte lists and synoptic bioassay data were located from Idaho or eastern Oregon. In addition, the only chronic test with sufficient data for inclusion was the *Hyalella azteca* 28-day test. While some surveys have been run in recent years using the *Chironomus dilutus* 20-day bioassay, there were less than 30 data points in total and only a few bioassay hits among those samples, which was not sufficient for development of SQVs. It appears that most project proponents are choosing to run the acute *Chironomus* test along with the chronic *Hyalella* test, thus limiting the availability of data for the chronic *Chironomus* test.

A complete list of surveys used for SQV development is provided in Appendix A.

2.1.2 Initial Data Screening

In assembling the data set, surveys, analytes, and individual data points were screened out if they did not meet certain initial data screening criteria, described below. Appendix B lists all the surveys, stations, and data that were screened out during assembly of the data set.

Completeness - Surveys and stations were screened out if they had an insufficient analyte list. Although it would be ideal for all stations to have the same analyte list when developing SQVs, this is not possible when using historical data sets. A minimum of semivolatiles and metals was selected as a general guideline for including a survey or station, consistent with other national criteria development efforts. Metals and semivolatiles both contribute significantly to toxicity in most contaminated sediment data sets, and if these minimum analytes were not available, toxicity would frequently occur in samples without adequate chemistry to explain it. This would lead to an unrealistically high number of false

negatives in the reliability analysis, based solely on the analyte list and not on the accuracy of the SQVs. For some surveys, different stations had varying analyte lists. In these surveys, only those stations with adequate analyte lists were retained.

Surveys were also screened out if insufficient information could be found to conduct chemistry and/or bioassay quality assurance evaluations. Both bioassay and chemistry data were subjected to quality assurance review at a level sufficient to support regulatory development and litigation, known as "QA2" (PTI 1989). Substantial efforts were made to obtain this information, including contacting the original clients, contractors, and laboratories. However, in some cases the data were too old, never had the required information, or could not be provided for a reasonable cost or within a reasonable timeframe.

Minimum Amount of Data - For development of SQVs, a minimum number of data points is required. To be as inclusive as possible, a minimum of 30 detected values was chosen as the lower limit for inclusion on the analyte list. Chemicals without enough detected data to calculate SQVs are listed in Appendix B. While these chemicals are not expected to be found in most projects, should they be important for a specific site, bioassay testing is recommended for evaluation of their potential toxicity.

Non-Toxicity - Analytes were also screened out for other reasons. Some analytes, such as iron, aluminum, and magnesium, were screened out because they are crustal elements and are naturally present in high concentrations, although some of these compounds can affect toxicity at certain sites. Certain conventional analytes, such as grain size parameters and acid-volatile sulfides, were screened out because they are physical parameters or other non-standard analytes. Others were derived quantities, such as dioxin toxicity equivalency quotients (TEQs). These analytes are listed in Appendix B.

Chemistry Quality Assurance - Individual chemical data were screened out based on qualifiers assigned during the quality assurance process. Data qualified as H, Q, X, or R (defined in Table 2-1 below) were not included in the analysis. Undetected data were also not included, as these data do not provide useful information for the purposes of developing SQVs. Data with these qualifiers were also excluded in Ecology's previous round of FPM calculations.

Table 2-1. Qualifier Definitions for Screened-Out Data

Qualifier	Definition
H	Holding time exceeded (conventionals)
Q	Questionable value
X	Less than 10% recovery
R	Rejected – failure to meet QA guidelines

Bioassay Quality Assurance - Some surveys and individual stations were screened out because of a low number of replicates in bioassays, below what is considered a minimum standard in modern freshwater protocols (ASTM 2005). Surveys or stations with less than five replicates were screened out. The freshwater ASTM protocols (ASTM 2005) recommend 8 replicates and require a minimum of 4 replicates in order to provide appropriate power under most circumstances. The minimum of 4 is mainly considered appropriate for less rigorous applications, such as trend analysis between years, and is fewer than the PSDDA marine bioassay standard of 5 replicates. The data sets remaining in the database after the above screening meet or exceed these minimum guidelines.

2.1.3 Normalization and Summing

To date, evaluations of the reliability of dry weight-normalized SQVs vs. organic carbon-normalized SQVs has shown that the dry weight values have equal or better reliability than the organic carbon-normalized values (PSEP 1988, Ecology 1997, SAIC and Avocet 2003). In addition, the use of organic carbon-normalized SQVs leads to implementation difficulties because it is difficult to understand and explain to the regulated community, and because it is inappropriate in some situations with large quantities of anthropogenically derived organic carbon. Consistent with regional dredging guidelines and all other SQVs calculated after the original marine Apparent Effects Thresholds (AETs), it was decided to calculate the SQVs on a dry weight normalized basis.

In the past, SQVs have been calculated both for individual polynuclear aromatic hydrocarbons (PAHs) and for summed dry weight values such as low molecular weight PAHs and high molecular weight PAHs. In recent years, there has been a trend toward using summed values of PAHs in the development of SQVs, as this may better reflect their mode of action and additive toxicity (Swartz et al., 1995; EPA 2000). A PAH workshop was held in June 2007 among the RSET agencies to discuss how best to handle petroleum toxicity in developing SQVs and bioaccumulative guidelines. The participants at this workshop selected the following approach for dealing with historical data sets.

Historical data should be evaluated on the basis of total PAHs, and total petroleum hydrocarbon (TPH) gasoline-, diesel-, and organic-range hydrocarbons. This could be accomplished by assembling one data set with total PAH values, and another data set with the TPH values. Normally, these two types of values should be considered as alternatives rather than being included in the same model run, as PAHs are a subset of TPH. Inclusion of both values in the same model run could theoretically produce unreliable results for one or both values, as they are not independent of one another. However, after multiple model runs it became apparent that TPH was the driving factor for petroleum toxicity rather than PAHs, although there were no TPH data for many stations. Therefore, both were retained in the model runs and the two together provided better reliability than either one alone.

Other sums used in the model runs included total dioxins/furans, total polychlorinated biphenyls (PCBs; sum of Aroclors), total chlordanes (sum of cis- and trans-chlordane, chlordane, alpha-chlordane, gamma-chlordane, cis- and trans-nonachlor, oxychlordane, heptachlor, and heptachlor epoxide), total endosulfans (alpha-endosulfan, beta-endosulfan, and endosulfan sulfate), total DDDs, total DDEs, and total DDTs (o,p' and p,p' isomers). Appendix B lists all of the constituents included in all sums, which were therefore not included as individual chemicals in the model runs.

The following summation rules were used for chemical classes:

- If all constituents were non-detects, the sum for that chemical class was treated in the same manner as non-detected individual chemicals, and excluded from model calculations.

- If some constituents were detected and others were non-detects, the non-detects were assigned a value of one-half the detected limit and summed with the other constituents.

- Unusually high non-detected values (e.g., due to interference) were not used; instead a value of one-half the standard detection limit for that analysis was used.

- Total PCBs calculated as a sum of Aroclors is an exception to the above summing rules. Aroclors that were undetected were assigned a value of zero. Because Aroclors are already a mixture of PCBs, and individual Aroclor products are frequently used in industrial processes in the absence of other Aroclor products, it cannot be assumed that non-detected Aroclor products are present.

Various methods of dealing with non-detected data were evaluated by the workgroup, including not including undetected constituents (i.e., setting their value to 0), using half the detection limit, or using statistical methods to estimate the true value. Using half the detection limit was selected for the following reasons:

- This approach is generally consistent with the approach outlined in Ecology's SMS regulations and with DEQ's standard practice. Because regulated parties will be required to calculate their sums in this manner, the SQVs should be calculated the same way so that comparisons are valid.

- It should reduce the variability and the error that would be associated with using zero for non-detected constituents of sums where most of the other constituents are detected.

- It is a simpler calculation procedure than available statistical methods, which would have to be developed, decided upon, and potentially applied differently depending on the distribution of each individual chemical sum.

2.1.4 Comparison to Control vs. Reference

Based on the results of SAIC and Avocet (2002), there appears to be no reliability advantage to using a comparison to reference rather than a comparison to control for this freshwater data set. Freshwater reference areas have not yet been standardized, and the variability of reference stations in the historical data set appears to overwhelm any theoretical advantage they may provide. In addition, many test stations do not have valid reference stations and would have to be excluded from the analysis if comparison to reference were used. Consequently, a comparison to control provides a much larger and more consistent data set to work with in calculating SQVs. Finally, all of the other national SQV sets that have been developed for freshwater have used a comparison to control. Therefore, it was decided to use comparison to control for derivation of SQVs.

This decision does not limit how individual regulatory programs may choose to interpret and use their bioassay data. It is anticipated that freshwater reference areas may be identified concurrently with this report as part of simultaneous RSET efforts, and once this process is completed it may be possible to use a comparison to reference for future updates of the SQVs. However, it is likely that the process may be more difficult than in the marine environment because of the more heterogeneous nature of freshwater environments, and there may not be valid reference areas for all freshwater sites.

2.1.5 Bioassay Tests and Endpoints

Five acute and chronic test endpoints had sufficient data to calculate SQVs:

- Chronic endpoints: *Hyalella azteca* 28-day growth and *Chironomus dilutus* 10-day growth,
- Acute endpoints: *Hyalella azteca* 10-day and 28-day mortality and *Chironomus dilutus* 10-day mortality.

While there were some *Chironomus dilutus* 20-day mortality and growth data collected, there were less than 30 data points total and only a few toxic stations, which is not sufficient for calculation of SQVs. Microtox was excluded after a lengthy evaluation process. Microtox protocols have changed sufficiently over the years that the data sets before and after the changes were not comparable, to the extent that attempts to combine these data sets resulted in poor reliability. There were insufficient data using the newer protocols to calculate SQVs. Therefore, it may be possible to calculate Microtox and *Chironomus dilutus* 20-day mortality and growth values in the future.

The first step in performing SQV calculations, once the data have been collected and screened, is the determination of whether adverse biological effects are observed in each sample (called a "hit" if observed and a "no-hit" if not observed). These biological effects levels may also be used to interpret the results of bioassay tests conducted to confirm or over-ride the chemical SQVs on an individual project.

The identification of adverse biological effects generally involves a statistical difference from the control or reference plus some threshold of effects, shown in Table 2-2 below. Quality assurance guidelines for control and reference samples are also shown. Derivation of the thresholds for each bioassay endpoint is discussed in detail following the table. In all cases, "statistically significant" means a statistical difference from a control sample at an alpha level of 0.05. Data transformations, selection of null hypotheses, and statistical testing procedures are identical to those currently in use by RSET for marine sediment data (Michelsen and Shaw 1996, Fox et al. 1998).

Table 2-2. Quality Assurance and Adverse Effects Levels for Biological Tests

Test	QA Control	QA Reference	SQS/SL1	CSL/SL2
Hyalella azteca 10-day mortality	C ≤ 20%[a]	R ≤ 25%	T – C > 15%	T – C > 25%
Hyalella azteca 28-day mortality	C ≤ 20%	R ≤ 30%	T – C > 10%	T – C > 25%
Hyalella azteca 28-day growth	CF ≥ 0.15 mg/ind	RF ≥ 0.15 mg/ind	T / C < 0.75	T / C < 0.6
Chironomus dilutus 10-day mortality	C ≤ 30%[a]	R ≤ 30%	T – C > 20%	T – C > 30%
Chironomus dilutus 10-day growth	CF ≥ 0.48 mg/ind	RF/CF ≥ 0.8	T / C < 0.8	T / C < 0.7

QA = Quality Assurance
SQS/SL1 = Sediment Quality Standard/Screening Level 1, CSL/SL2 = Cleanup Screening Level/Screening Level 2
C = Control, CF = Control Final, R = Reference, RF = Reference Final, T = Test Sample
[a] These control mortality limits are currently in the process of being reviewed by ASTM and may be lowered in the next few years (Ingersoll et al. 2008)

Hyalella azteca 10-day mortality bioassay

- **SQS/SL1 mortality:** A hit requires a statistically significant difference from control, and a relative increase in mortality of > 15% (test – control > 15%).
- **CSL/SL2 mortality:** A hit requires a statistically significant difference from control, and a relative increase in mortality of > 25% (test – control > 25%).

The ASTM protocols originally established a control performance standard of 20% mortality, although in practice, the mean mortality observed in the control samples in round robin testing was approximately 10%. Recently, it has been recommended that the control performance standard be modified to 15% mortality (Ingersoll et al. 2008). Given this, the maximum mortality that would be observed at the SQS/SL1 level would be 30-35%, and would often be less, and the maximum mortality that would be observed at the CSL/SL2 level would be 40-45%, and would often be less. This SQS/SL1 level would be very similar in practice to the marine SQS/SL1 level of 30% absolute mortality.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 5 to 24%, with a mean of 11%. Therefore, a detectable difference could be observed at levels as low as 10-15% mortality, ranging in the worst case up to about 35% mortality, depending on the performance of the control samples and the degree of variability in the test replicates. In practice these thresholds should be observable nearly all of the time, with the minimum detectable difference occasionally exceeding the SQS/SL1 numeric threshold, but not likely exceeding the SL2 numeric threshold.

***Hyalella azteca* 28-day mortality bioassay**

- **SQS/SL1 mortality:** A hit requires a statistically significant difference from control, and a relative decrease in mortality of > 10% (test – control > 10%).
- **CSL/SL2 mortality:** A hit requires a statistically significant difference from control, and a relative increase in mortality of > 25% (test – control > 25%).

The ASTM protocols establish a control performance standard of 20% mortality, and the results of round robin testing reported that > 90% of laboratories were able to meet that standard. Given this, the maximum mortality that would be observed at the SQS/SL1 level would be 30%, and would often be less, and the maximum mortality that would be observed at the CSL/SL2 level would be 45%, and would often be less. This approach sets the same policy goals as the acute mortality test, but gives a little more latitude in the reference performance standard for the challenges of running a longer test.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 2 to 20%, with a mean of 8%. Therefore, a detectable difference could be observed at levels as low as 15% mortality, ranging in the worst case up to about 35% mortality, depending on the performance of the control samples and the degree of variability in the test replicates. In practice these endpoints should be observable most of the time, with the minimum detectable difference at times exceeding the SQS/SL1 numeric threshold, but not likely exceeding the SL2 numeric threshold.

***Hyalella azteca* 28-day growth bioassay**

- **SQS/SL1 growth:** A hit requires a statistically significant difference from control, and a relative decrease in weight of > 25% (test/control < 75%).
- **CSL/SL2 growth:** A hit requires a statistically significant difference from control, and a relative decrease in weight of > 40% (test/control < 60%).

The SQS/SL1 and CSL/SL2 endpoints are based largely on the minimum detectable differences reported in ASTM round robin studies, since little additional information exists on which to base recommendations. The mean minimum detectable difference in weight in round robin studies was

approximately 25%, with a range from 16 to 50%. Balancing these considerations are literature studies suggesting that reductions in growth of as little as 20-30% can cause significant reproductive effects and other physiological changes in aquatic species, including *Chironomus dilutus* and *Mytilus galloprovincialis* (ASTM 2005, Kagley et al. 1995, Widdows & Donkin 1992). The recommended endpoints above are a compromise between statistical reality and environmental policy objectives. The round robin studies suggest that the numeric level corresponding to the SQS/SL1 should be observable about half the time, and the numeric level corresponding to the CSL/SL2 should be observable about 80% of the time.

It should be noted that the length measurement is substantially less variable than the weight measurement in assessing growth effects, and would be preferable to use in the future for that reason. However, most laboratories have not yet installed the equipment that would allow for automation of this endpoint, and historic data are expressed in weight (ASTM 2005). The suggested control and reference performance standard, based on the draft ASTM protocol, is greater than or equal to 0.15 mg mean individual biomass at time final.

***Chironomus dilutus* 10-day mortality bioassay**

- **SQS/SL1 mortality:** A hit requires a statistically significant difference from control, and a relative decrease in mortality of > 20% (test – control > 20%).
- **CSL/SL2 mortality:** A hit requires a statistically significant difference from control, and a relative increase in mortality of > 30% (test – control > 30%).

The ASTM protocols establish a control performance standard of 30% mortality, although in practice, the mean mortality observed in the control samples in round robin testing was approximately 8%, with a range of 1-19%. Recently, it has been recommended that this be reduced to 20% (Ingersoll et al. 2008). Given this, the maximum mortality that would be observed at the SQS/SL1 level would be 40%, and would usually be less, and the maximum mortality that would be observed at the CSL/SL2 level would be 50%, and would usually be less.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 2 to 12%, with a mean of 8% (the mortality endpoint did not appear to be as sensitive to noise/variability as either the 10-day *Hyalella* mortality endpoint or the 10-day *Chironomus* growth endpoint). Therefore, a detectable difference could be observed at levels as low as 15% mortality, ranging in the worst case up to about 30% mortality, depending on the performance of the control samples and the degree of variability in the test replicates. In practice these numeric thresholds should be observable nearly all of the time.

***Chironomus dilutus* 10-day growth bioassay**

- **SQS/SL1 growth:** A hit requires a statistically significant difference from control, and a relative decrease in weight of > 20% (test/control < 80%).
- **CSL/SL2 growth:** A hit requires a statistically significant difference from control, and a relative decrease in weight of > 30% (test/control < 70%).

The SQS/SL1 and CSL/SL2 endpoints are based largely on the minimum detectable differences reported in ASTM round robin studies. The mean minimum detectable difference in weight in round robin studies

was approximately 11%, with a range from 5 to 24%. This allows for more protective SQS/SL1 and CSL/SL2 levels than for either of the chronic growth tests. The round robin studies suggest that the numeric level corresponding to the SQS/SL1 should be observable well over half of the time, and the CSL/SL2 levels should be observable nearly all of the time. The numeric levels chosen span the range of growth rates associated with adverse reproductive or physiological effects in the literature, as discussed above.

The control performance standards established for the 10-day test are equal to or greater than 0.48 mg mean individual biomass at time final, and the recommended reference performance standard is at least 80% of the control.

2.1.6 ANOVA Analyte Screening

A second screening of the data set was conducted to remove chemicals that are not apparently associated with toxicity in this data set. This was accomplished by comparing the hit and no-hit distributions to determine if they were statistically different using an ANOVA comparison, with various p values ≤ 0.1, 0.05, 0.005, and 0.0005 to show increasing degrees of association with toxicity. Experience with application of the FPM has shown that chemicals with hit and no-hit distributions that are not statistically different do not affect the reliability of the SQVs developed using that data set. This was verified in some early runs on the Portland Harbor project, as well as recent projects conducted for Ecology (Avocet 2003), ODEQ (1999), San Francisco Bay, and Los Angeles Harbor.

Detailed results of the ANOVA screening evaluations, which were conducted separately for each chemical, effects level, and endpoint combination, are provide in Appendix B. Because the same chemicals did not always contribute to toxicity in all tests and endpoints, the list of chemicals included in the modeling for each endpoint is different. These differences could be due to a variety of factors, including differences in the response of test organisms or endpoints to the chemicals, and differences in the underlying data sets for each test endpoint.

Certain chemicals had no relationship to benthic toxicity for any of the hit/no-hit definitions or endpoints. These included Aldrin, dioxins/furans, gamma-hexachlorocyclohexane, hexachlorobenzene, hexachloroethane, methoxychlor, retene, and total endosulfans. These chemicals were not included in the subsequent model runs and should not be considered chemicals of concern for benthic toxicity at the range of concentrations observed in this database. However, many of these chemicals may still exhibit toxicity to wildlife or human health through bioaccumulative exposure routes and should be evaluated accordingly. Other chemicals were screened out for some endpoints, but nevertheless have final SQVs because they were related to toxicity for other endpoints. Chemicals screened out as a result of the ANOVA screening are listed in Appendix B, along with the underlying ANOVA matrices.

2.2 SQV Calculation - Floating Percentile Method

The basic concept behind the FPM is to select an optimal percentile of the data set that provides a low false negative rate and then adjust individual chemical concentrations upward until false positive rates are decreased to their lowest possible level while retaining the same low false negative rate. As shown in Figure 2-1, the y-axis is the percentile of each chemical's overall distribution and is not linearly related to toxicity. The green vertical line shows the concentration range within which toxicity does not occur,

and the red vertical line shows the range within which toxicity occurs. These ranges may overlap due to site-specific or sample-specific variations in bioavailability or toxicity.

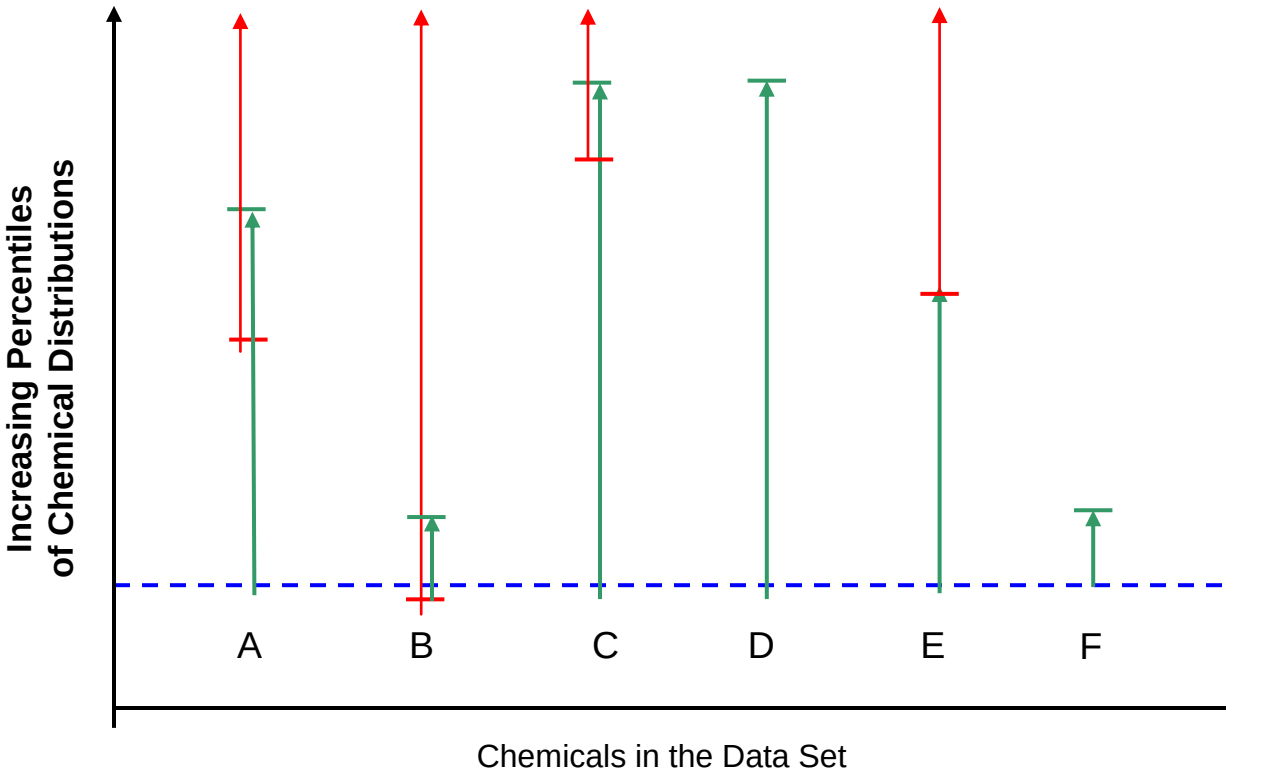
A constant percentile of the distribution that results in a low false negative rate is initially selected for all chemicals, represented by the blue dashed line. The difference between this constant percentile and the lower end of the toxicity range for each chemical is the area between the blue line and the red bar, and this is the source of most of the false positive errors.

The second step is to determine which chemicals are associated with false positive errors in the data set and adjust those concentrations upward until the lower end of their toxicity ranges are reached (red bar). Above this point, false negatives will begin to increase. Above the red bar, both false negatives and false positives may occur, as is shown for Chemicals A, B, and C. This region is the range of concentrations over which sample-specific bioavailability plays an important role in toxicity, and therefore hit and no-hit samples are mixed together, causing both types of errors.

In Figure 2-1, Chemical B's concentration cannot be raised at all because it is already within its toxic concentration range. In any data set, a few chemicals will already be at a toxic level, giving rise to the low percentage of false negatives that the blue line represents. Some chemicals may show a sharper toxicity threshold (e.g., Chemical E). Others may not appear to be related to toxicity in the data set at all (e.g., Chemicals D and F). These chemical concentrations can be raised to their maximum percentile without any observed increase in toxicity. However, it may be safer in practice to raise them only to the point where false positives no longer occur (represented by the green bar) or to a similar endpoint such as AETs.

Once each chemical has been individually adjusted upward to the lower end of its toxicity range, the false positives will have been significantly reduced while retaining the same low false negative rate. Most chemicals should be at or near their actual toxicity range, rather than at a level arbitrarily assigned by a fixed percentile. In this manner, optimized site-specific SQVs can be developed for a number of different target false negative rates, allowing the trade-offs between false negatives and false positive to be evaluated and a final set of SQVs to be selected.

Figure 2-1. Floating Percentile Method



Legend:

Fixed percentile for all chemicals

Region within which false positives occur

Toxicity range within and above which false negatives occur

In summary, the steps required to calculate SQVs using this approach include:

- Compile and screen synoptic chemistry/bioassay data.
- Select toxicity tests and endpoints.
- Assign hit/no-hit status for each station/endpoint combination.
- Develop chemical distributions.
- Select a range of target false negative rates and identify associated optimal percentile values.
- Adjust percentiles for individual chemicals upward to reduce false positives.

Optimization of chemical concentrations occurs through an iterative automated step using an Excel macro. The Excel macro uses the following approach to conduct the optimization:

1. An appropriate incremental increase for testing is selected for each analyte based on that analyte's complete concentration range (e.g., 1/10 of the difference between the highest and lowest concentration).

2. The number of false positives contributed by each individual analyte is calculated, and the chemical contributing the most false positives is selected to begin the optimization procedure.

3. The concentration for that analyte is increased by the chosen increment.

4. After each incremental increase, false negative and false positive rates are recalculated for the entire SQV set.

5. If the false negative rate increases, the chemical concentration is adjusted back down to its previous level and that chemical is "locked in" at that level.

6. If the false positive rate is reduced to zero, the chemical concentration is locked in at that level.

7. If either of the above two conditions is met, or if the number of false positives for that chemical has been reduced below that of another chemical, the macro moves on to the chemical with the current highest number of false positives. If none of these criteria are met, the macro raises the concentration by another increment and repeats steps 4-7.

8. Incremental increases and recalculations continue until every chemical has reached its toxicity threshold or a level at which it has no more false positives.

Through this process, it is possible to identify those analytes having the greatest influence on toxicity in the data set (those whose concentrations cannot be increased without increasing false negatives), and those chemicals having little or no influence on toxicity in the data set (those that can be increased to their highest concentrations with no effect on error rates).

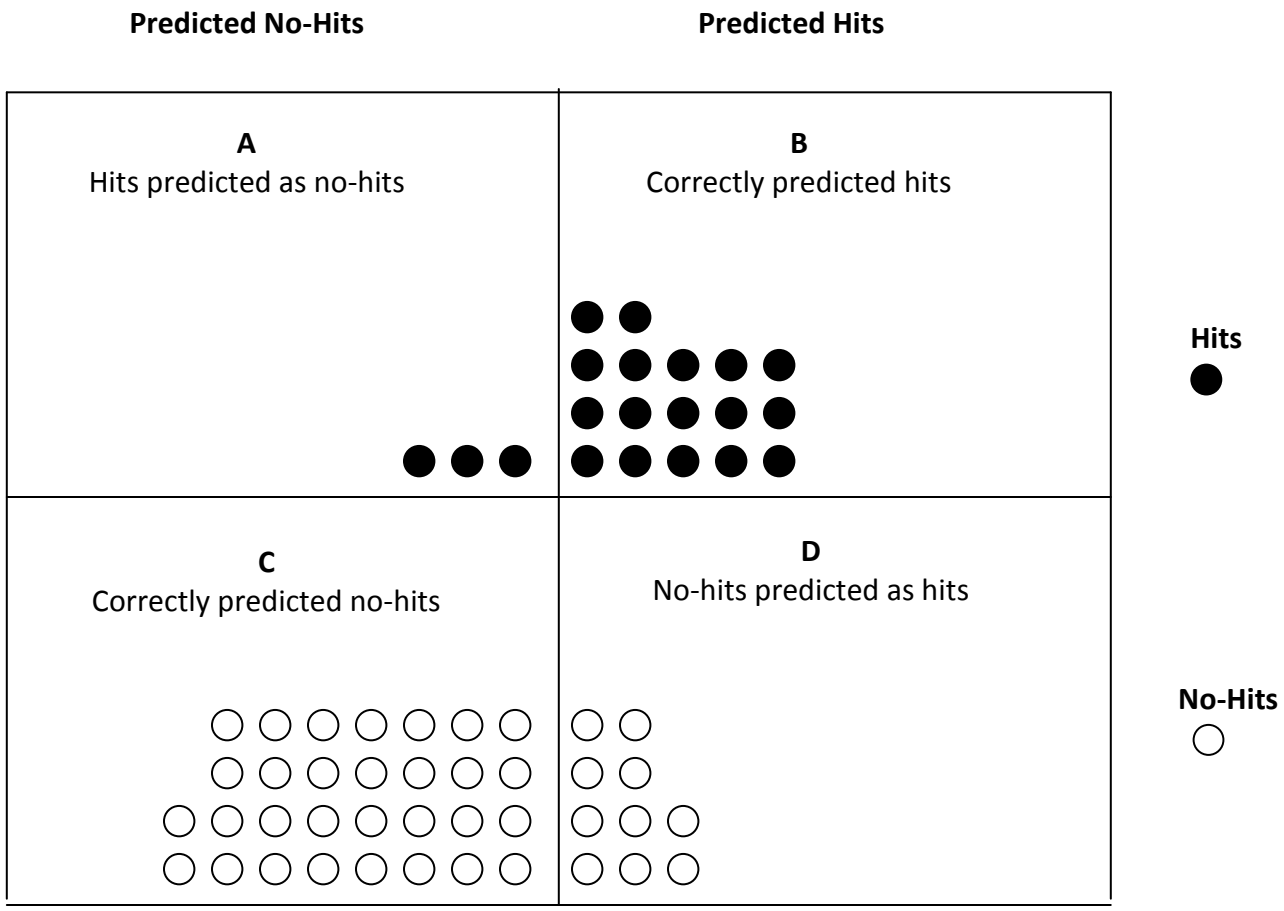
The spreadsheets used to develop the SQVs also provide a test area, where candidate SQV sets may be adjusted and finalized, and the results of each change tested with respect to all of the reliability parameters (this area also allows the operator to enter any criteria set of their choice and test its reliability against the regional data set). For transparency, the workgroup chose not to conduct any operator finalization, and instead to work directly with the results of the automated procedure.

2.3 Reliability Analysis

Reliability analysis was conducted following the derivation of the SQVs. The measures of reliability that were used are defined and illustrated graphically below:

- **False Negatives:** hits incorrectly predicted as no-hits/total number of hits
- **False Positives:** no-hits incorrectly predicted as hits/total number of no-hits
- **Sensitivity:** hits correctly predicted/total number of hits (100% - % false negatives)
- **Efficiency:** no-hits correctly predicted/total number of no-hits (100% - % false positives)
- **Predicted Hit Reliability:** correctly predicted hits/total predicted hits
- **Predicted No-Hit Reliability:** correctly predicted no-hits/total predicted no-hits
- **Overall Reliability:** correct predictions/total stations

Figure 2-2. Reliability Measures



Sensitivity = B / (A + B)
False Negatives = A / (A + B)

Efficiency = C / (C + D)
False Positives = D / (C + D)

Predicted-Hit Reliability = B / (B + D)
Predicted-No-Hit Reliability = C / (A + C)

Overall Reliability = (B + C) / (A + B + C + D)

False positives and false negatives are the primary measure of predictive errors in the reliability assessment. Each of the other reliability values is related to them in some way.

While the performance of any given data set cannot be determined in advance, the workgroup agreed on a set of reliability goals that would guide the selection of the final SQVs, shown in Table 2-3. The goals were set based on two factors: 1) the levels of error the agencies were collectively comfortable with in making regulatory decisions, and 2) the levels of reliability that were considered reasonably achievable based on previous results of the FPM model. It should be noted that most or all existing SQV sets have much higher levels of error, and that the reliability that can be achieved is constrained by environmental and laboratory variability in the bioavailability of chemicals, sensitivity of organisms, etc.

Table 2-3. Reliability Goals for Proposed Freshwater SQVs

	SQS/SL1 (%)		**CSL/SL2 (%)**	
	SEF[a]	**Goal**	**SEF[a]**	**Goal**
Sensitivity	84	80 – 90	85	75 - 85
Efficiency	75	70 – 80	75	75 - 85
Predicted Hit Reliability	88	70 – 80	77	75 - 85
Predicted No-Hit Reliability	67	80 – 90	84	75 - 85

[a] Actual value achieved for interim SEF freshwater SQVs

2.4 Exploratory Model Runs

Exploratory model runs were conducted for a variety of scenarios to explore data relationships and provide information on the best possible ways to work with the data set. The following separate model runs were conducted:

- **Bioassay Endpoints –** The model was run separately for each individual bioassay endpoint at both SQS/SL1 and CSL/SL2 effects levels. This allows an evaluation of the bioassay endpoints with respect to each other – for example, which ones behave similarly, which chemical groups each responds to, and which endpoints are most sensitive and reliable. This information informs the selection of bioassay endpoints for use in setting the final SQVs as well as for use at individual sites, and may also point out areas where further bioassay development is needed.

- **Pooled Endpoints –** As noted above, the model was run for each bioassay endpoint individually, the results of which would be combined later to develop the draft SQS/SL1 and CSL/SL2 SQVs. In addition, a wide variety of "pooled" endpoints were run. A pooled endpoint combines the results of multiple toxicity tests into a single hit/no-hit determination, which is then used to run the model. The ways in which multiple bioassays were combined included:

    - All stations were included and a hit on any one or more bioassays was defined as an overall hit
    - Only stations with 2 or more or 3 or more bioassays were included, and a hit on any one or more bioassays was an overall hit
    - Only stations with both acute and chronic data were included, and any one or more hits was an overall hit
    - All stations were included, and 2 or more hits at the SQS/SL1 level or a single hit at CSL/SL2 level was defined as a hit.

In some cases, stations with only one bioassay and a no-hit result were considered indeterminate, since additional bioassays might have shown a hit.

- **PAHs vs. TPH.** The model was run with one data set in which total PAHs were summed on a dry weight basis, and also on a smaller data set containing TPH data for gasoline-, diesel-, and residual-range hydrocarbons. In this manner, both total PAHs and TPH values were calculated (from separate data sets). This approach allows a comparison of the reliability of these measurement endpoints in predicting petroleum-related toxicity. In addition, the model was run with both endpoints included.

- **East-side vs. West-side.** The model was run for the entire data set, as well as separately using data east of the Cascades and west of the Cascades. This approach reflects the widely differing geochemistry, industries, and analytes associated with these two areas and allowed evaluation of whether different SQVs would be appropriate for these georegions.

- **N-Qualified Data.** There was a fair amount of chemistry data for pesticides that was N-qualified, meaning that the identity of the chemical was indeterminate, likely due to degraded spectra. Considerable debate occurred about whether or not these data should be included, with valid arguments on both sides. The model was run both with and without these data to inform the discussion.

- **Blank-Correction.** It was determined during the quality assurance review that the data sets had not all been blank-corrected in the same manner. After addressing this issue by re-qualifying the data sets in a consistent manner, the models were run again to demonstrate an improvement in reliability.

2.5 Final Model Runs

Based on the exploratory model runs, the following decisions were made and are reflected in the final model runs:

- The reliability of individual endpoints varied significantly, particularly at the lower SQS/SL1 level. Pooled model runs, rather than averaging or otherwise reflecting all the endpoints at a station, tended to have reliability as poor as the least reliable endpoint included in the analysis. Therefore, only the individual endpoints were used (similar to AETs) to select SQVs.

- Microtox data were not used in the final runs. There were issues with the quality of the older data and not enough data using the newer protocols to calculate reliable SQVs.

- Total PAHs, as well as TPH-diesel and TPH-residual, were included in the final model runs. The reliability was best when both were included. Of the two alone, TPH was more predictive; however, TPH data were missing for many data sets, leading to improved reliability when both were included.

- East- and west-side data were combined into a single data set. The reliability was best when both were combined. It is possible that reliability could be further improved by removing mining-related watersheds from the overall data set.

- N-qualified data were included in the data set, as this improved reliability for these chemicals.
- For stations with detected concentrations in the blanks, consistently revising qualifiers according to the EPA Contract Laboratory Protocols improved reliability.

2.6 Decision Framework and Selection of SQVs

2.6.1 Regulatory Considerations

Two effects levels were developed for each bioassay endpoint, one corresponding to SQS/SL1 and one corresponding to CSL/SL2. SQS/SL1 is based on minimum detectable difference from control or reference and represents a no observable effects level, and CSL/SL2 adds an allowable degree of minor adverse effects.

For dredging projects in the RSET program, the SL1 serves as a threshold above which biological testing would be required. SL2 serves as an advisory threshold above which the agencies believe biological testing is likely to fail, but applicants could still conduct biological testing in hopes of passing.

In the Washington State sediment cleanup program, the SQS serves as the long-term goal for sediments of the state, and the lower end of the range within which cleanup standards for a site can be selected. The CSL serves as the level above which cleanup sites are designated, and also serves as the upper end of the range within which cleanup standards for a site may be selected, based on balancing environmental protectiveness, cost, and technical feasibility. Thus, a cleanup standard for any given site may be set within a range of allowable adverse effects, from no effects to minor adverse effects, depending on site-specific considerations. This regulatory framework is the same for both freshwater and marine standards, and thus the approach used to develop the freshwater SQVs was as similar as possible to the marine standards in terms of overall structure, level of protectiveness, and biological effects interpretive guidelines.

As with the marine SQVs, the draft freshwater SQVs were specifically selected to provide an appropriate balance of sensitivity and efficiency (i.e., balancing false negatives and false positives) on a per-sample basis, while retaining a low enough false negative rate to ensure that contaminated sites would be identified given the amount of data typically available for site identification purposes. To ensure that the SQVs are adequately protective, they will be applied within a regulatory framework that includes the option of conducting bioassays as a confirmatory or override step, or simultaneously with chemical analyses. The same suite of bioassays and interpretive endpoints used to develop the SQVs will also be used to interpret the bioassay results. These were selected to reflect the range of species and life history stages of a benthic community and their use for both purposes ensures consistency and maximizes the reliability of the SQV predictions.

In general, the freshwater SQVs were developed to protect populations of benthic communities in sediments, rather than individual species, given the wide natural variation in species abundance and richness seasonally and from year to year, especially in freshwater systems. NOAA and USF&W assisted the RSET workgroup in determining whether the SQS/SL1 was protective of individual ESA-listed benthic species, and determined that there were no listed benthic species in WA, OR, or ID that were present in areas where dredging or cleanup is likely to be conducted. Therefore, lower values to protect individual ESA-listed benthic species did not need to be developed.

2.6.2 Technical Approach

As noted above, the model was run for each individual bioassay endpoint separately, at two effects levels corresponding to SQS/SL1 and CSL/SL2. This approach is desirable because it preserves information about bioassay endpoint sensitivity and reliability, the relationships between bioassay endpoints, and the relationships between chemicals and toxicity for different endpoints. In addition, it reduces potential problems with pooling bioassay toxicity results ahead of time that are associated with variations in historical data, including bioassay endpoints at each station, chemical analytes at each station, number and variability of replicates, etc.

Based on the results of modeling, differences in the SQVs between bioassays proved to be much larger than differences between the SQS/SL1 and CSL/SL2 levels for any one bioassay endpoint. Therefore, the values for all the bioassay endpoints and effects levels were combined into a single distribution from which the SQS/SL1 and CSL/SL2 would be selected. This distribution reflects the range of SQVs from the lowest no-effects level to the highest minor effects level. The following method is recommended for setting the final SQVs:

- **SQS/SL1 –** Select the lowest value from among the distribution
- **CSL/SL2** – Select the next highest significantly different value (i.e., not within laboratory replicate guidelines).

This approach provides conservative values by remaining at the low end of the no-adverse-effects to minor-adverse-effects distribution, while still providing a degree of distance between the two levels for regulatory flexibility in decision-making.

3.0 RESULTS

3.1 Final Data Set

The numbers of stations for each bioassay endpoint in the final data set are shown in Table 3-1, comprising 5 distinct sample/test combinations. Tables 3-1 and 3-2 do not include samples that failed quality assurance requirements.

Table 3-1. Bioassays and Endpoints in Final Data Set

Test	No. of Samples
Hyalella azteca 10-day mortality	366
Hyalella azteca 28-day mortality	312
Hyalella azteca 28-day growth	79
Chironomus dilutus 10-day mortality	568
Chironomus dilutus 10-day growth	525

These samples are associated with 648 stations having various combinations of bioassays at each station, of which 583 are from west of the Cascades (WA and OR) and 65 are from east of the Cascades (WA). Table 3-2 shows the number and percentage of stations associated with biological hits for each bioassay and effects level.

Table 3-2. Biological Hits at Each Effects Level

Effects Level	SQS/SL1[a]	CSL/SL2[a]
Hyalella azteca 10-day mortality	89 (24%)	52 (14%)
Hyalella azteca 28-day mortality	47 (15%)	27 (7%)
Hyalella azteca 28-day growth	26 (33%)	12 (15%)
Chironomus dilutus 10-day mortality	85 (15%)	41 (7%)
Chironomus dilutus 10-day growth	65 (12%)	49 (9%)

[a] See Table 2-2 for SQS/SL1 and CSL/SL2 definitions

3.2 Reliability Assessment

Using the FPM, a variety of exploratory and final model runs were conducted as described in Sections 2.4 and 2.5 to obtain results that met the original reliability goals set by the workgroup. Tables 3-3 and 3-4 show the reliability results for six different choices of false negative rates at the SQS/SL1 and the

CSL/SL2 levels. Dark blue rows meet the reliability goals selected by the workgroup. Light blue rows are within 5% and are considered borderline. Yellow rows do not meet the reliability goals. As can be seen in the tables below, each bioassay endpoint at each effects level had at least one row that met the reliability goals. However, reliability was considerably better at the CSL/SL2 level, farther from the minimum detectable difference.

The cross-hatched box in each of the tables below indicates the row that was selected by the workgroup for derivation of SQVs. In each case, the selected rows met the reliability goals established by the workgroup. Therefore, the FPM values developed were considered appropriately sensitive, efficient, and reliable, in accordance with the workgroup's reliability goals.

In addition, reliability tests were run for other existing SQV sets to determine their predictiveness with this data set, including:

- For comparison with SQS/SL1 levels: Effects Range Low (ERL), Threshold Effects Levels (TELs), Threshold Effects Concentrations (TECs), and Lower Effects Levels (LELs).
- For comparison with CSL/SL2 levels: Effects Range Median (ERM), Probable Effects Levels (PELs), Probable Effects Concentrations (PECs), and Severe Effects Levels (SELs).

These results are shown underneath each table below for ease of comparison. In comparison with other existing SQV sets, the following can be seen:

- At the SQS/SL1 level, the false positives for the existing SQV sets are typically in the 75-95% range, 2-3 times higher than those of the FPM values at an equivalent false negative level. Overall reliability of the existing SQV sets is low, in the 15-45% range, compared to 70-95% for the selected FPM values. None of the existing SQV sets had a combination of sensitivity, efficiency, and overall reliability that fell within the workgroup's reliability goals for any test, in contrast to the FPM values.

- At the CSL/SL2 level, the existing SQV sets still had at least twice the false positive rate of the FPM values, but often had twice the false negative rates as well. Overall reliability was typically 10-30% lower than the FPM values. In only two cases did an existing SQV set come within 5% of the reliability goals set by the workgroup.

Therefore, the FPM values represent a significant improvement in reliability over the available SQVs at both upper and lower effects levels.

Table 3-3. Reliability of the FPM Results and Existing SQV Sets at the SQS/SL1 Level

Legend for all tables:

Does not meet reliability goals
Borderline reliability (within 5% of goals)
Meets reliability goals
Meets reliability goals; selected for development of SQVs

FPM FN Percentiles – False negative target for the modeling run

SQVs – Existing Sediment Quality Guidelines:

ERL - Effects Range Low, TEL - Threshold Effects Levels, TEC - Threshold Effects Concentrations, LEL - Lower Effects Levels, ERM - Effects Range Median, PEL - Probable Effects Levels, PEC - Probable Effects Concentrations, and SEL - Severe Effects Levels

a. Chironomus 10-day growth

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.6	44.8	95.4	55.2	23.1	98.8	60.2
10	9.2	35.9	90.8	64.1	26.3	98.0	67.4
15	13.8	31.7	86.2	68.3	27.7	97.2	70.5
20	20.0	17.0	80.0	83.0	40.0	96.7	82.7
25	24.6	19.6	75.4	80.4	35.3	95.9	79.8
30	29.2	13.5	70.8	86.5	42.6	95.4	84.6

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERL	6.2	85.9	93.8	14.1	13.4	94.2	24.0
TEL	4.6	91.3	95.4	8.7	12.9	93.0	19.4
TEC	7.7	79.6	92.3	20.4	14.1	94.9	29.3
LEL	9.2	88.3	90.8	11.7	12.7	90.0	21.5

b. Chironomus 10-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.7	40.8	95.3	59.2	29.1	98.6	64.6
10	9.4	33.1	90.6	66.9	32.5	97.6	70.4
15	14.1	26.5	85.9	73.5	36.3	96.7	75.4
20	20.0	21.3	80.0	78.7	39.8	95.7	78.9
25	24.7	19.7	75.3	80.3	40.3	94.9	79.6
30	29.4	16.6	70.6	83.4	42.9	94.2	81.5

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERL	9.2	86.7	90.8	13.3	27.9	79.7	34.2
TEL	5.9	91.3	94.1	8.7	27.5	80.0	31.7
TEC	11.1	79.5	88.9	20.5	29.2	83.3	38.9
LEL	6.5	87.5	93.5	12.5	28.3	83.9	34.3

c. Hyalella 10-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.5	59.2	95.5	40.8	34.1	96.6	54.1
10	9.0	48.0	91.0	52.0	37.9	94.7	61.5
15	14.6	35.7	85.4	64.3	43.4	93.2	69.4
20	19.1	32.5	80.9	67.5	44.4	91.7	70.8
25	24.7	28.9	75.3	71.1	45.6	90.0	72.1
30	29.2	27.1	70.8	72.9	45.7	88.6	72.4

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERL	2.8	87.5	97.2	12.5	32.0	91.4	37.7
TEL	2.8	88.3	97.2	11.7	31.8	90.9	37.2
TEC	8.3	74.7	91.7	25.3	34.2	87.8	45.1
LEL	4.6	80.9	95.4	19.1	33.3	90.7	41.8

d. Hyalella 28-day growth

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	3.8	52.8	96.2	47.2	47.2	96.2	63.3
10	7.7	52.8	92.3	47.2	46.2	92.6	62.0
15	11.5	43.4	88.5	56.6	50.0	90.9	67.1
20	19.2	18.9	80.8	81.1	67.7	89.6	81.0
25	23.1	17.0	76.9	83.0	69.0	88.0	81.0
30	26.9	17.0	73.1	83.0	67.9	86.3	79.7

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERL	13.8	83.3	86.2	16.7	29.0	75.5	36.4
TEL	3.4	93.7	96.6	6.3	28.9	82.4	31.8
TEC	13.8	84.6	86.2	15.4	28.6	73.9	35.4
LEL	3.4	94.1	96.6	5.9	28.8	81.3	31.5

e. Hyalella 28-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.3	48.3	95.7	51.7	26.0	98.6	58.3
10	8.5	35.8	91.5	64.2	31.2	97.7	68.3
15	14.9	23.8	85.1	76.2	38.8	96.7	77.6
20	19.1	12.5	80.9	87.5	53.5	96.3	86.5
25	23.4	11.3	76.6	88.7	54.5	95.5	86.9
30	29.8	9.1	70.2	90.9	57.9	94.5	87.8

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERL	10.6	83.4	89.4	16.6	16.0	89.8	27.6
TEL	4.3	94.3	95.7	5.7	15.3	88.2	19.2
TEC	10.6	84.5	89.4	15.5	15.8	89.1	26.6
LEL	6.4	95.1	93.6	4.9	14.9	81.3	18.3

Table 3-4. Reliability of the FPM Results and Existing SQV Sets at the CSL/SL2 Level

a. Chironomus 10-day growth

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.1	40.8	95.9	59.2	19.5	99.3	62.7
10	8.2	34.7	91.8	65.3	21.4	98.7	67.8
15	14.3	22.3	85.7	77.7	28.4	98.1	78.5
20	18.4	12.4	81.6	87.6	40.4	97.9	87.0
25	24.5	13.7	75.5	86.3	36.3	97.2	85.3
30	28.6	12.8	71.4	87.2	36.5	96.7	85.7

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERM	14.3	41.4	85.7	58.6	17.6	97.6	61.1
PEL	18.4	42.0	81.6	58.0	16.7	96.8	60.2
PEC	30.6	29.8	69.4	70.2	19.3	95.7	70.1
SEL	40.8	23.1	59.2	76.9	20.9	94.8	75.2

b. Chironomus 10-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	4.9	30.0	95.1	70.0	19.8	99.5	71.8
10	9.8	26.0	90.2	74.0	21.3	99.0	75.2
15	14.6	22.0	85.4	78.0	23.2	98.6	78.5
20	19.5	18.0	80.5	82.0	25.8	98.2	81.9
25	24.4	12.9	75.6	87.1	31.3	97.9	86.3
30	29.3	9.3	70.7	90.7	37.2	97.6	89.3

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERM	28.4	43.5	71.6	56.5	18.0	93.7	58.3
PEL	28.4	44.5	71.6	55.5	17.7	93.6	57.4
PEC	40.3	31.7	59.7	68.3	20.1	92.7	67.3
SEL	50.7	24.6	49.3	75.4	21.2	91.7	72.4

c. Hyalella 10-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	3.8	60.5	96.2	39.5	20.8	98.4	47.5
10	9.6	56.4	90.4	43.6	21.0	96.5	50.3
15	13.5	45.2	86.5	54.8	24.1	96.1	59.3
20	19.2	28.0	80.8	72.0	32.3	95.8	73.2
25	25.0	24.8	75.0	75.2	33.3	94.8	75.1
30	28.8	20.7	71.2	79.3	36.3	94.3	78.1

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERM	30.8	43.6	69.2	56.4	20.8	91.7	58.2
PEL	30.8	40.4	69.2	59.6	22.1	92.1	60.9
PEC	46.2	28.7	53.8	71.3	23.7	90.3	68.9
SEL	51.9	19.4	48.1	80.6	29.1	90.4	76.0

d. Hyalella 28-day growth

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	0.0	29.9	100.0	70.1	37.5	100.0	74.7
10	8.3	16.4	91.7	83.6	50.0	98.2	84.8
15	8.3	16.4	91.7	83.6	50.0	98.2	84.8
20	16.7	13.4	83.3	86.6	52.6	96.7	86.1
25	25.0	11.9	75.0	88.1	52.9	95.2	86.1
30	25.0	11.9	75.0	88.1	52.9	95.2	86.1

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERM	50.0	45.9	50.0	54.1	6.3	94.6	53.9
PEL	50.0	49.3	50.0	50.7	5.9	94.2	50.6
PEC	61.1	35.9	38.9	64.1	6.3	94.4	62.7
SEL	55.6	30.0	44.4	70.0	8.4	95.3	68.5

e. Hyalella 28-day mortality

FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	3.7	11.6	96.3	88.4	44.1	99.6	89.1
10	7.4	7.7	92.6	92.3	53.2	99.2	92.3
15	14.8	4.6	85.2	95.4	63.9	98.6	94.6
20	18.5	4.2	81.5	95.8	64.7	98.2	94.6
25	22.2	3.5	77.8	96.5	67.7	97.9	94.9
30	29.6	1.8	70.4	98.2	79.2	97.2	95.8

SQVs	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
ERM	37.0	45.3	63.0	54.7	11.6	94.0	55.4
PEL	25.9	47.7	74.1	52.3	12.8	95.5	54.2
PEC	33.3	34.0	66.7	66.0	15.7	95.4	66.0
SEL	29.6	28.1	70.4	71.9	19.2	96.2	71.8

3.3 Sediment Quality Guidelines

Further conservativism was employed in selecting the draft SQVs. As noted in the methods section, the FPM values for individual bioassay endpoints differed greatly among the bioassays, but often were similar for the same bioassay between the two effects levels. As a result, all the values were combined for each chemical into a single distribution that represented values from the no effects to the minor adverse effects range. Each chemical had between 4 and 10 values, depending on the number of bioassay endpoints for which an FPM value could be developed for that chemical. Tables 3-5 and 3-6 show the FPM values for each endpoint, based on the rows selected above. “Greater than” signs (>) indicate that the toxicity value for that chemical and bioassay is greater than any of the concentrations in the database, and the maximum concentration is shown in the table.

Based on the modeling results, several chemicals were not found to be toxic to benthic organisms at the range of concentrations found in the data set; all of their FPM values were “greater than” values. These chemicals included butyl benzyl phthalate, dimethyl phthalate, and total chlordanes (in addition to the chemicals previously excluded in earlier steps, such as ANOVA screening). These chemicals were eliminated from the benthic SQV set, but should still be evaluated for bioaccumulation effects to human health and wildlife.

The remaining concentrations were ordered from lowest to highest for each chemical, as shown in Table 3-7. To develop SQVs, the lowest concentration was selected as the SQS/SL1, and the next highest significantly different concentration was selected as the CSL/SL2. Some professional judgment was used in determining whether values were significantly different, based on quality assurance guidelines for replicate analyses. The resulting proposed SQV values are also shown in Table 3-7.

For some chemicals, only an SQS/SL1 could be established; the remaining concentrations were all “greater than” concentrations. This suggests that, for these chemicals, only low levels of effects are

observed within the concentration range included in this data set. Higher levels of effects may be observed above the "greater than" value. Therefore, that value has been included as the CSL for site managers' information. At levels above those observed in this data set, bioassays should be run to identify the presence or absence of higher levels of adverse effects.

It should be noted that these are only draft SQV recommendations, based on the many selections and method assumptions outlined in this report. Alternative choices could be made that would change the SQVs. In addition, implementing agencies and programs may choose to adopt all or only some of the SQVs shown in the table, depending on their program priorities. The final decisions on how to proceed will be made by Ecology, Oregon DEQ, RSET, and the other agencies and programs that may choose to use these values, following appropriate public review and comment.

Table 3-5. Floating Percentile Model Values at the SQS/SL1 Level

Analyte	CH10G	CH10M	HY10M	HY28G	HY28M
Conventional Pollutants (mg/kg)					
Ammonia	> 780		> 780		230
Total sulfides	39	540	920		61
Metals (mg/kg)					
Antimony	42		0.3	42	12
Arsenic	120	120	200	14	16
Cadmium	6.3	2.1	13	23	5.4
Chromium	88	220		72	82
Copper	1600	1900		400	> 1900
Lead	360	> 1400	> 1300	> 1400	> 1400
Mercury	3	0.8		0.66	0.87
Nickel	110	> 590	360	26	100
Selenium	> 20			11	> 20
Silver	0.58	0.64			1.7
Zinc	> 14000		> 4200	3200	3200
Organic Chemicals (µg/kg)					
4-Methylphenol	> 6300	2000	2400		260
Benzoic acid		2900	3800		
beta-Hexachlorocyclohexane	7.2	11			11
bis(2-Ethylhexyl)phthalate	> 440000		500		> 440000
Butylbenzyl phthalate	> 2800	> 2800			> 2800
Carbazole	1400	1100	2900		30000
Dibenzofuran	> 7200	680	3800		680
Dibutyltin	910	910			> 910
Dieldrin	4.9	4.9			22
Dimethyl phthalate	> 580	> 580			
Di-n-butyl phthalate	380	450			1000
Di-n-octyl phthalate	> 1100		39		
Endrin ketone	8.5	8.5			8.5
Monobutyltin	540	540			> 540
Pentachlorophenol	> 1200	> 1200	1200		> 320
Phenol	> 770	210	250		210
Tetrabutyltin	97	97			> 97
Total Aroclors	3100	3400	110		3400
Total Chlordanes	> 670	> 670			> 670
Total DDDs	860	2500	310		2500
Total DDEs	900	900	21	> 5.7	900
Total DDTs	> 13000	100			8100
Total PAHs	30000	45000	17000		330000
Tributyltin	9300	320			> 9300
Bulk Petroleum Hydrocarbons (mg/kg)					
TPH-Diesel	540	340	1700		1700
TPH-Residual	4400	3600	> 8400		10000

SQS/SL1 = Sediment Quality Standard/Screening Level 1
CH10G = *Chironomus* 10-day growth, CH10M = *Chironomus* 10-day mortality,
HY10M = *Hyalella* 10-day mortality, HY28G = *Hyalella* 28-day growth, HY28M = *Hyalella* 28-day mortality
> "greater than" value indicates that the toxic level is unknown, but above the concentration shown

Table 3-6. Floating Percentile Model Values at the CSL/SL2 Level

Analyte	CH10G	CH10M	HY10M	HY28G	HY28M
Conventional Pollutants (mg/kg)					
Ammonia	> 780		> 780		300
Total sulfides	340	360	920		340
Metals (mg/kg)					
Antimony	42		0.3	42	> 63
Arsenic	120	180	200	14	16
Cadmium	6.3	2.1	13	> 23	> 23
Chromium	220	> 350	> 350	72	220
Copper	1600	1900	> 11000	1200	> 1900
Lead	360	> 1400	> 1300	> 1400	> 1400
Mercury	0.66	0.8	0.8	> 0.87	0.87
Nickel	110	> 590	360	> 27	> 100
Selenium	> 20			11	> 20
Silver	4.1	4.1	4.1		1.7
Zinc	> 14000		> 4200	3200	> 14000
Organic Chemicals (µg/kg)					
4-Methylphenol	> 6300	2000	2400		260
Benzoic acid		4100	3800		
beta-Hexachlorocyclohexane	11	20			11
bis(2-Ethylhexyl)phthalate	> 440000		22000		> 440000
Butylbenzyl phthalate	> 2800	> 2800	> 1500		> 2800
Carbazole	1400	2500	2900		30000
Dibenzofuran	200	7200	3800		7200
Dibutyltin	910	910	130000		> 910
Dieldrin	4.9	9.3			22
Dimethyl phthalate	> 580	> 580	> 580		
Di-n-butyl phthalate	> 1800	450	> 1700		1000
Di-n-octyl phthalate	> 1100		39		
Endrin ketone	8.5	8.5			8.5
Monobutyltin	540	540	> 4800		> 540
Pentachlorophenol	> 1200	> 1200	1200		> 320
Phenol	> 770	210	250		120
Tetrabutyltin	97	97			> 97
Total Aroclors	3400	3400	2500		3400
Total Chlordanes	> 670	> 670	> 180		> 670
Total DDDs	> 3000	2500	310		2500
Total DDEs	900	33	> 44	> 5.7	900
Total DDTs	> 13000	8100	> 140		8100
Total PAHs	17000	77000	33000		1700000
Tributyltin	9300	320	48		> 9300
Bulk Petroleum Hydrocarbons (mg/kg)					
TPH-Diesel	510	390	2100		1700
TPH-Residual	4400	8400	> 8400		10000

CSL/SL2 = Cleanup Screening Level/Screening Level 2
CH10G = *Chironomus* 10-day growth, CH10M = *Chironomus* 10-day mortality,
HY10M = *Hyalella* 10-day mortality, HY28G = *Hyalella* 28-day growth, HY28M = *Hyalella* 28-day mortality
> "greater than" value indicates that the toxic level is unknown, but above the concentration shown

Table 3-7. Selection of Recommended Sediment Quality Guidelines

Analyte	Distribution of Floating Percentile Model Values[a]										SL1/SQS[b]	SL2/CSL[c]
Conventional Pollutants (mg/kg)												
Ammonia	230	300	> 780	> 780	> 780	> 780					230	300
Total sulfides	39	61	340	340	360	540	920	920			39	61
Metals (mg/kg)												
Antimony	0.3	0.3	12	42	42	42	42	> 63			0.3	12
Arsenic	14	14	16	16	120	120	120	180	200	200	14	120
Cadmium	2.1	2.1	5.4	6.3	6.3	13	13	> 23	> 23	> 23	2.1	5.4
Chromium	72	72	82	88	220	220	220	> 350	> 350		72	82
Copper	400	1200	1600	1600	1900	1900	> 1900	> 1900	> 11000		400	1200
Lead	360	360	> 1300	> 1300	> 1400	> 1400	> 1400	> 1400	> 1400	> 1400	360	> 1300
Mercury	0.66	0.66	0.8	0.8	0.8	0.87	0.87	> 0.87	3.04		0.66	0.8
Nickel	26	> 27	> 100	> 100	110	110	360	360	> 590	> 590	26	110
Selenium	11	11	> 20	> 20	> 20	> 20					11	> 20
Silver	0.58	0.64	1.7	1.7	4.1	4.1					0.58	1.7
Zinc	3200	3200	3200	> 4200	> 4200	> 14400	> 14400	> 14400			3200	> 4200
Organic Chemicals (µg/kg)												
4-Methylphenol	260	260	2000	2000	2400	2400	> 6300	> 6300			260	2000
Benzoic acid	2900	3800	3800	4100							2900	3800
beta-Hexachlorocyclohexane	7.2	11	11	11	11	20					7.2	11
bis(2-Ethylhexyl)phthalate	500	22000	> 440000	> 440000	> 440000	> 440000					500	22000
Carbazole	1100	1400	1400	2500	2900	2900	30000	30000			1100	1400
Dibenzofuran	200	680	680	3800	3800	7200	7200	> 7200			200	680
Dibutyltin	910	910	910	910	> 910	> 910	130000	130000			910	130000
Dieldrin	4.9	4.9	4.9	9.3	22	22					4.9	9.3
Di-n-butyl phthalate	380	450	450	1000	1000	> 1700	> 1800				380	450
Di-n-octyl phthalate	39	39	> 1100	> 1100							39	> 1100
Endrin ketone	8.5	8.5	8.5	8.5	8.5	8.5					8.5	**
Monobutyltin	540	540	540	540	> 540	> 540	> 4800				540	> 4800
Pentachlorophenol	> 320	> 320	1200	1200	> 1200	> 1200	> 1200	> 1200			1200	> 1200
Phenol	120	210	210	210	250	250	> 770	> 770			120	210
Tetrabutyltin	97	97	97	97	> 97	> 97					97	> 97
Total Aroclors	110	2500	3100	3400	3400	3400	3400	3400			110	2500
Total DDDs	310	310	860	2500	2500	2500	2500	> 3000			310	860
Total DDEs	> 5.7	> 5.7	21	33	> 44	900	900	900	900	900	21	900
Total DDTs	100	> 140	8100	8100	8100	> 13000	> 13000				100	8100
Total PAHs	17000	17000	30000	33000	45000	77000	330000	1700000			17000	30000
Tributyltin	48	320	320	9300	9300	> 9300	> 9300				48	320

Analyte	Distribution of Floating Percentile Model Values[a]								SL1/SQS[b]	SL2/CSL[c]
Bulk Petroleum Hydrocarbons (mg/kg)										
TPH-Diesel	340	390	510	540	1700	1700	1700	2100	340	510
TPH-Residual	3600	4400	4400	8400	> 8400	> 8400	10000	10000	3600	8400

SQS/SL1 = Sediment Quality Standard/Screening Level 1, CSL/SL2 = Cleanup Screening Level/Screening Level 2

> "greater than" value indicates that the toxic level is unknown, but above the concentration shown

[a] The model concentrations for each of the bioassay endpoints and each of the effects levels shown in Tables 3-5 and 3-6 are shown here as a distribution of values from lowest to highest.

[b] The value selected as the SQS/SL1 was the lowest of all the values in the distribution.

[c] The value selected as the CSL/SL2 was the second lowest value in the distribution; the second lowest value was defined as the next-highest value that was significantly different from the SQS/SL1 value. Minor differences within quality assurance guidelines for replicate analyses were not considered "significantly different."

4.0 CONCLUSIONS

In summary, the following observations and conclusions can be drawn:

- **Synoptic Bioassay/Chemistry Data Set.** The freshwater data set is considerably larger and more diverse in terms of both chemistry and bioassays than it was in 2003, and has been improved from a quality assurance standpoint. The current database allows calculation of FPM values for five acute and chronic endpoints.

- **Geographic Representativeness.** Data sets were collected from western Washington and Oregon and from eastern Washington. No data were identified in eastern Oregon or Idaho that included synoptic bioassay and chemistry data. The data set encompasses a wide variety of different types of environments, including large and small lakes on both sides of the Cascades, large rivers on both sides of the Cascades such as the Duwamish, Willamette, Columbia, and Spokane Rivers, and small streams.

- **Sensitivity, Efficiency, and Reliability.** Use of the floating percentile method resulted in endpoint-specific SQVs with a sensitivity of 75-80%, an efficiency of 65-95%, and an overall reliability of 70-85%, depending on the specific endpoint and effects level. Because the lowest of these values is proposed as the SQS/SL1 and the next highest value as the CSL/SL2, the resulting SQVs are from the low end of the distribution and will have higher overall sensitivity and somewhat lower efficiency than the individual endpoint-specific values.

- **Comparison to Existing SQVs.** The FPM values represent a substantial improvement in efficiency and overall reliability for comparable false negative rates. In addition, at the higher effects levels, the FPM values are also much more sensitive than the existing SQV sets.

- **Recommended SQVs.** Based on the conclusions above and an approach developed by the interagency workgroup for combining the individual endpoint values, SQVs for both the SQS/SL1 and the CSL/SL2 levels are recommended. The method provides the opportunity for revision of these values if alternative policy choices regarding sensitivity and efficiency are made during the agency and public review process.

- **Benthic Toxicity Only.** It should be reiterated that these values were developed to protect against toxicity to the benthic community only. They are not protective of bioaccumulative effects to humans, wildlife, or fish. However, based on a review and consultation with NOAA and USF&W regarding ESA species in WA, OR, and ID, they are expected to be protective of ESA-listed benthic species.

- **Additional Information for Site Managers.** Additional information for site managers is included in Appendix B, including a list of chemicals that were screened out and the reasons for doing so, and how to evaluate chemicals that may not have recommended SQVs.

5.0 REFERENCES

ASTM. 2005. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. ASTM E1706-05. American Society for Testing and Materials, West Conshohocken, PA.

Brunelle, H, C Mach, K Parrett. 2003. Evaluating Polycyclic Aromatic Hydrocarbons Ecological Threshold Concentrations for Sediment Using Logistic Regression Modeling. Poster presentation at Pacific NW SETAC conference, April 17-19, 2003, Fort Worden, WA. Prepared by Ecology and Environment for Oregon Department of Environmental Quality, Portland OR.

Deneer, JW, TL Sinnege, W Seinen, and JLM Hermens. 1988. The joint acute toxicity to *Daphnia magna* of industrial organic chemicals at low concentrations. *Aquat. Toxicol.* 12:33-38.

DMEF. 1998. Dredged Material Evaluation Framework Lower Columbia River Management Area. U.S. Army Corps of Engineers, Portland and Seattle Districts; EPA Region 10, Seattle, WA; Washington Department of Ecology, Olympia, WA; Oregon Department of Environmental Quality, Portland, OR; Washington Department of Natural Resources, Olympia, WA.

Ecology. 1997. Creation and Analysis of Freshwater Sediment Quality Values in Washington State. Washington Department of Ecology, Environmental Investigations and Laboratory Services Program, Olympia WA.

EPA. 2000. Equilibrium Partitioning Sediment Guidelines (ESGs) for the Protection of Benthic Organisms: PAH Mixtures. U.S. Environmental Protection Agency, Office of Science and Technology and Office of Research and Development.

Field, LJ, SB Norton, DD MacDonald, CG Severn, CG Ingersoll. 2003. Predicting Toxicity from Sediment Chemistry using Logistic Regression Models: Regional and Site-Specific Applications. Presentation at Pacific NW SETAC conference, April 17-19, 2003, Fort Worden, WA. National Oceanic and Atmospheric Administration, Coastal Protection and Restoration Division, Seattle, WA.

Fox, D.F., D.A. Gustafson, and T.C. Shaw. 1998. Biostat Software for the Analysis of DMMP/SMS Bioassay Data. DMMP Clarification Paper, SMS Technical Information Memorandum. Seattle District Corps of Engineers, Seattle, WA.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York, NY.

Hermens, J, H Canton, P Janssen, R de Jong. 1984. Quantitative structure-activity relationships and toxicity studies of mixtures of chemicals with anaesthetic potency: Acute lethal and sublethal toxicity to *Daphnia magna. Aquat. Toxicol.* 5:143-154.

Hermens, J, E Brockhuyzen, H Canton, R Wegman. 1985a. Quantitative structure activity relationships and mixture toxicity studies of alcohols and chlorohydrocarbons: Effects on growth of *Daphnia magna. Environ. Toxicol. Chem.* 4:273-279.

Hermens, J, P Leeuwangh, A Musch. 1985b. Joint toxicity of mixtures of groups of organic aquatic pollutants to the guppy (*Poecilia reticulata*). *Ecotox. Environ. Safety* 9:321-326.

Ingersoll, CG, CD Ivey, NE Kemble, DR Mount, J Field, DD MacDonald, D Smorong. Compilation of control performance data for laboratories conducting whole-sediment toxicity tests with the amphipod Hyalella azteca and the midge Chironomus dilutus (formerly C. tentans). SETAC National Meeting, 2008.

Michelsen, T.C. 1999. Error rate minimization techniques for calculating sediment quality guidelines. Presentation at SETAC North America conference, 2003, Philadelphia, PA. Avocet Consulting, Kenmore, WA.

Michelsen, T.C. and T.C. Shaw. 1996. Statistical Evaluation of Bioassay Results. PSDDA Clarification Paper, SMS Technical Information Memorandum. Washington Department of Ecology, Olympia, WA, and Seattle District Corps of Engineers, Seattle, WA.

PSEP. 1988. 1988 Update and Evaluation of Puget Sound AET. U.S. Environmental Protection Agency, Puget Sound Estuary Program, Seattle, WA.

PTI Environmental Services. 1989. Puget Sound Dredged Disposal Analysis Guidance Manual: Data Quality Evaluation of Proposed Material Disposal Projects (QA-2). Prepared for Washington Department of Ecology Sediment Management Unit, Olympia WA.

SAIC and Avocet. 2002. Development of Freshwater Sediment Quality Values in Washington State, Phase I Final Report. Prepared by SAIC, Bothell, WA and Avocet Consulting, Kenmore, WA for the Washington Department of Ecology, Olympia, WA.

SAIC and Avocet. 2003. Development of Freshwater Sediment Quality Values in Washington State, Phase II Final Report. Prepared by SAIC, Bothell, WA and Avocet Consulting, Kenmore, WA for the Washington Department of Ecology, Olympia, WA.

Sokal, R.R., and F.J. Rohlf. 1981. Biometry. Second Edition. W.H. Freeman and Company, San Francisco, CA.

Swartz, RC, DW Schults, RJ Ozretich, JO Lamberson, FA Cole, TH DeWitt, MS Redmond, and SP Ferraro. 1995. ∑PAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environmental Toxicology and Chemistry* 14(11):1977-1987.

APPENDIX A. LIST OF SURVEYS

Survey	CHR10G	CHR10M	HYA10M	MCTX	CHR20G	CHR20M	HYA28G	HYA28M	Description
BOISECAS	0	0	4	0	0	0	0	0	Class II Inspection of the Boise Cascade Pulp and Paper Mill Wallula Washington, WA Dept. of Ecology EILS, 1993
CARGIL01	0	3	3	0	0	0	0	0	Cargill Irving Elevator Terminal, Cargill Irving, 2001
CBSLOUGH	0	0	20	0	0	0	0	0	Columbia Slough Sediment Analyses and Remediation Project, Phase 1 Report, Dames & Moore for City of Portland, 1991
CEDARIV	0	0	5	5	0	0	0	0	Sediment Sampling and Analysis Report Cedar River Delta Sediments, Golder Assts. for City of Renton, 1992
FWDMMP05	26	26	26	0	0	0	0	0	Sediment Characterization Report, Lower Willamette River Federal Navigational Channel, Corps of Engineers, 2005
FWJSLK04	8	8	8	0	0	0	0	0	Johnson Lake Site Investigation Report, Arcadis for Owens-Brockway Glass Container, Inc., 2004
FWLKUN01	5	4	4	0	0	0	0	0	Lake Union Sediment Study, King County DNR, 2001
FWPHBR04	227	233	0	0	0	0	0	233	Portland Harbor Remedial Investigation Round 2 Data, Lower Willamette Group, 2004
FWSPOR00	0	0	0	8	7	8	8	8	Chemical Analysis and Toxicity Testing of Spokane River Sediments Collected in October 2000, WA Dept. of Ecology EAP, 2001
FWTEKX07	3	3	3	0	0	0	0	0	Tektronix Site Remedial Investigation, Phase III, Windward Environmental, 2007
FWUPCR05	50	50	0	0	0	0	50	50	Upper Columbia River Site CERCLA RI/FS, CH2M Hill for US EPA Region 10, 2005
FWWRSD04	21	21	21	0	21	21	21	21	Willamette River Federal Navigation Channel O&M Sediment Characterization Report, Corps of Engineers, 2004
LCBWRS93	0	0	15	0	0	0	0	0	Lower Columbia River Backwater Reconnaissance Survey, TetraTech for Lower Columbia River Bi-State Program, 1994
LKUNDRDK	0	0	4	0	0	0	0	0	Sediment Monitoring Program Results Lake Union Drydock Company, Hart Crowser, 1992
LKUNION	0	0	9	0	0	0	0	0	Survey of Contaminants in Lake Union and Adjoining Waters, WA Dept. of Ecology EILS, 1989
LKWA00	0	28	28	27	0	0	0	0	Lake Washington Baseline Sediment Study, King County, 2000
LUUCSO00	0	6	6	6	0	0	0	0	Lake Union University Regulator CSO Post Separation Study, King County, 2000
MBCREOS3	43	43	43	0	0	0	0	0	McCormick & Baxter RD Phase I Sediment Survey, Oregon DEQ, 2002
MBCREOS4	17	18	18	0	0	0	0	0	McCormick & Baxter RD Phase II Sediment Survey, Oregon DEQ, 2002
PPTLDT24	4	4	4	0	0	0	0	0	Sediment Characterization Study, Marine Terminal 2 Berths 203-206 and Marine Terminal 4 Berth 416, Hart Crowser for Port of Portland, 1999
PSYD&M97	0	0	3	0	0	0	0	0	Portland Shipyard Environmental Audit, Dames & Moore for Cascade General, 1998
PSYSEA98	55	55	55	55	0	0	0	0	Portland Shipyard Sediment Investigation Data Report, Striplin Env. Assts. for Port of Portland, 1998
QUEBAX1	0	0	4	0	0	0	0	0	Distribution and Significance of PAHs in Lake Washington Sediments Adjacent to Quendall Terminals, WA Dept. of Ecology EILS, 1991
QUEBAX3	0	0	3	0	0	0	0	0	Results of Sediment Sampling in the JH Baxter Cove Lake Washington, WA Dept. of Ecology EILS, 1992
ROSSIS99	11	11	11	0	0	0	0	0	Ross Island Facility Site Investigation, Hart Crowser for Port of Portland, 2000
SALIII97	22	22	22	22	0	0	0	0	Salmon Bay Results of Phase III Sampling, WA Dept of Ecology EAP, 2000
SEACOM94	0	0	3	3	0	0	0	0	Sediment Sampling Report Seattle Commons Parcel C Seattle, Washington, 1994
SPOKNR94	0	0	3	3	0	0	0	0	Spokane River PCB Study, WA Dept of Ecology EILS, 1994
TOSCO99	2	2	2	0	0	0	0	0	TOSCO Portland Terminal, 1999 Sediment Sampling Results, Portland District Corps of Engineers, 1999
TRI-STAR	0	0	3	0	0	0	0	0	Tri-Star Marine NPDES Sediment Monitoring, Beak Consultants, 1997
WEYLONG	0	0	3	0	0	0	0	0	Class II Inspection of Weyerhaeuser Longview Pulp and Paper Mill, WA Dept. of Ecology EILS, 1991
WILREF02	3	3	3	0	0	0	0	0	Willamette Reference Survey, Hart Crowser for the Portland District Corps of Engineers, 2002
WLRPT498	18	18	18	0	0	0	0	0	Terminal 4 Slip 3 Sediment Investigation, Hart Crowser for Port of Portland, 1998
WRD&M98	0	0	2	0	0	0	0	0	Portland Shipyard Environmental Audit, Dames & Moore for Cascade General, 1998
TOTAL	**515**	**558**	**356**	**129**	**28**	**29**	**79**	**312**	

APPENDIX B

DATA SCREENING

Section 2.1.2 describes the data screening that was conducted during assembly of the data set and prior to conducting the initial model runs. This appendix provides details of the surveys and data that were screened out.

Surveys and Stations

The following surveys and stations were identified but were screened out, for the reasons given (survey codes are SEDQUAL codes and indicate surveys already entered into SEDQUAL/EIM).

Two early data sets from the McCormick & Baxter Creosoting Company RI/FS (**MBCREOS1** and **MBCREOS2**) were removed from the data set when it was determined that the logistic regression models using the *Hyalella azteca* results for these data sets were significantly different from the rest of the *H. azteca* data sets. These studies were conducted in the 1990-1991 timeframe, and unlike more recent studies, the *H. azteca* organisms were collected locally and may have had a different sensitivity to contaminants. Although for some time there has been a general sense that the early McCormick & Baxter results were unusual, this was recently confirmed in a more rigorous manner by both NOAA (Field et al. 2003) and the Oregon Department of Environmental Quality (Brunelle et al. 2003).

Similarly, the 28-day *Hyalella azteca* growth data from the Portland Harbor RI were ultimately screened out, after much discussion among the agencies. These bioassay data did not show a correlation to any toxic chemicals in the study area, and had poor reliability in the modeling results. Removal of these data dramatically increased the usability and reliability of the overall *Hyalella azteca* 28-day growth data set. The EPA site managers, the SQV workgroup, and the Lower Willamette Group concurred with this decision. However, all other Portland Harbor bioassay data, including the *Hyalella azteca* 28-day mortality data, were retained.

In addition, some surveys and individual stations were screened out because of a low number of replicates in bioassays, below what is considered a minimum standard in modern freshwater protocols (ASTM 2000). Surveys or stations with less than five replicates were screened out, including:

- **LAKROO92 (all 18 stations) –** 7-day Hyalella, 3 replicates.
- **LSAMM99 (all 16 stations)** – Microtox®, 2 replicates
- **MARCO90 (1 station) –** 10-day Hyalella, 3 replicates.
- **QUEBAX2 (all 4 stations) –** 14-day Hyalella, 4 replicates.
- **SIMILK00 (all 4 stations) –** 10-day Hyalella, 4 replicates.
- **TRISTAR (all 3 stations)** – Microtox®, 3 replicates.
- **UNIMAR2 (all 9 stations)** – 14-day Hyalella, 3 replicates.

Surveys and stations were also screened out if they had an insufficient analyte list. A minimum of semivolatiles and metals was selected as a general guideline for including a survey or station, consistent with other national criteria development efforts. For some surveys, different stations had varying

analyte lists. In these surveys, only those stations with adequate analyte lists were retained. The surveys and stations screened out included:

- **COLALU94 (all 6 stations) –** Only conventionals.
- **LKROOS92 (2, 8, 10, 11, 15, 17, 19, 61, 71)** – 6 metals and TOC.
- **LKROOS01 (all 10 stations) –** 6 metals plus conventionals.
- **SIMILK00 (all 4 stations) –** metals and conventionals, no organics.
- **STEILLK2 (all 4 stations) –** metals and conventionals, no organics.
- **QUEBAX2 (all 4 stations) –** PAHs and conventionals, no metals.
- **Pope & Talbot Wood Treating Facility, St. Helens, OR** – insufficient chemistry
- **Zidell 2007** – Study still underway, data incomplete
- **Fifteen Mile Creek, OR** – no chemistry other than oxyfluorfen
- **Lower Clear Water River, WA** – no chemistry co-located with bioassays
- **Spokane River 2003, WA** – conventionals and a few metals
- **Mill Creek, WA** – conventionals and a few metals
- **Upper Columbia River 2001, WA** – conventionals and a few metals

Additional data sets were eliminated because insufficient information could be found to conduct QA2 review for either chemistry data or bioassay data or both; or other key information such as lat/longs or the SAP was missing:

- **Modoc Lumber, OR** – missing QA/QC information, SAP, and station locations
- **Weyerhaeuser Klamath Falls** – missing QA/QC information, station locations, and bioassay SAP
- **Pacific Carbide** – missing QA/QC for chemistry, bioassay failed QA/QC review
- **Tri-Met Merlo Garage, OR** – missing SAP, station locations, QA/QC
- **Nichols Boat Works, OR** – missing chemistry QA/QC

Thirteen samples were also deleted from a 2001 Lake Union survey because the percent solids in these samples ranged between 6-26%. This is very low for sediment samples and suggests that these samples were actually floc-like watery material, which would not be representative of typical sediments. Five remaining samples with percent solids > 45% were retained in the data set.

Analytes

Analytes were also screened out, for a variety of reasons. The following analytes are not toxic chemicals, and were screened from the initial data set:

- Grain size parameters
- Total organic carbon
- Total solids
- Acid volatile sulfides
- Derived parameters: Dioxin/furan TEQs (individual and summed dioxin and furan concentrations were retained)

Certain crustal elements were also removed from the dataset; these parameters are analyzed as part of standard metals suites, but are not known to be toxic at concentrations typically encountered in sediments:

- Aluminum
- Calcium
- Iron
- Magnesium
- Manganese
- Potassium
- Sodium

Certain chemicals were detected less than 30 times in the data set; these chemicals were also screened out as being unlikely to significantly influence toxicity in this large a data set. These chemicals will rarely be encountered, but if they should be encountered at high concentrations at a specific site or hot spot area, bioassay analyses should be conducted to evaluate their toxicity.

Table B-1. Rarely Detected Analytes

Chemical Analytes	No. Detections
1,2,3,4-Tetrahydronaphthalene	1
1,2,3-Trichloropropane	1
1,2,4-Trichlorobenzene	6
1,2-Dichlorobenzene	12
1,2-Dichloroethane	1
1,4-Dichlorobenzene	26
2,3,4,5-Tetrachlorophenol	5
2,3,4,6-Tetrachlorophenol	1
2,4,5-Trichlorophenol	25
2,4-D	6
2,4-DB	1
2,4-Dichlorophenol	2
2,4-Dimethylphenol	4
2,4-Dinitrotoluene	1
2-Chloronaphthalene	1
2-Chlorophenol	1
2-Methylphenol	8
4-Chloro-3-methylphenol	5
4-Nitroaniline	1
4-Stigmasten-3-one	1
7,10,13-Hexadecatrienoicacid	1
9-Hexadecenoicacid	2
Abietic acid	4
Aniline	12
Benzene	19
Benzyl alcohol	28
Bis(2-chloroethyl) ether	2

Caprolactam	1
Carbon disulfide	15
Chlorobenzene	17
Chloroform	21
Chloromethane	1
cis-1,2-Dichloroethene	2
Dehydroabietic acid	3
Dichloromethane	8
Diethyl phthalate	17
Endrin aldehyde	12
Ethylbenzene	16
gamma-Sitosterol	3
Hexachlorobutadiene	32*
Isophorone	3
Isopimaric acid	4
m,p-Xylene	20
MCPA	2
MCPP	2
Methyl iodide	1
Methyl tert-butyl ether	7
Methylethyl ketone	27
Mirex	7
N-Nitrosodiphenylamine	4
o-Xylene	29
Perylene	8
Phytol	3
Pimaric acid	4
Pristane	7
Sandaracopimaric Acid	1
Styrene	22
Thallium	13
Toluene	16
Trichloroethene	6
Xylenes	2

* This analyte has > 30 detections in the entire data set, but < 30 detections for any one bioassay endpoint.

Several analytes had enough detected values to be included, but not enough “hit” values for calculation of SQVs (< 10). These chemicals included alpha-, delta-, and gamma-hexachlorocyclohexane, Endrin, beryllium, and vanadium. These analytes were excluded from the modeling runs.

A number of chemicals were summed into groups and the individual analytes removed from the data set. The toxicity of these chemicals is additive or synergistic within their groups and is best represented by the group as a whole. Individual SQVs do not need to be established for these constituents, as their toxicity is represented by their group. The groups and their constituents are listed below:

- **DDD isomers:** o,p'-DDD, p,p'-DDD
- **DDE isomers:** o,p'-DDE, p,p'-DDE
- **DDT isomers:** o,p'-DDT, p,p'-DDT
- **Dioxins/Furans:** Total heptachlorodibenzofurans, total heptachlorodibenzo-p-dioxins, total hexachlorodibenzofurans, total hexachlorodibenzo-p-dioxins, octachlorodibenzofuran, octachlorodibenzo-p-dioxin, total pentachlorodibenzofurans, total pentachlorodibenzo-p-dioxins, total tetrachlorodibenzofurans, total tetrachlorodibenzo-p-dioxins
- **Total Chlordanes:** alpha-chlordane, chlordane, cis-chlordane, cis-nonachlor, gamma-chlordane, heptachlor, heptachlor epoxide, oxychlordane, trans-chlordane, trans-nonachlor
- **Total Endosulfans:** alpha-endosulfan, beta-endosulfan, endosulfan sulfate
- **Total PAHs:** 1-methylnaphthalene, 2-methylnaphthalene, acenaphthene, acenaphthylene, anthracene, benz(a)anthracene, benzo(a)pyrene, benzo(b)fluoranthene, benzo(ghi)perylene, benzo(k)fluoranthene, chrysene, dibenz(ah)anthracene, fluoranthene, fluorene, indeno(123-cd)pyrene, naphthalene, phenanthrene, pyrene, total benzofluoranthenes (b+k+j)
- **Total PCBs:** Aroclors 1016, 1221, 1242, 1248, 1254, 1260, 1268 (no congener data were available)

ANOVA Screening

One of the first steps of the model runs is to evaluate which chemicals show a relationship to toxicity in the data set, for each chemical and each endpoint (Table B-2). This evaluation is described in Section 2.1.6. As a result of this evaluation, it was determined that the following chemicals showed no relationship to toxicity for any of the endpoints, and these chemicals were not retained for further modeling:

- Aldrin
- dioxins/furans
- gamma-hexachlorocyclohexane
- hexachlorobenzene
- hexachloroethane
- methoxychlor
- retene
- total endosulfans

These chemicals have not been demonstrated to be toxic to the benthic community at sediment concentrations historically observed in the environment, and SQVs do not need to be set for them.

In addition to these chemicals, some chemicals were not related to toxicity for some tests and endpoints. These were screened out of modeling runs for these endpoints, but overall SQVs may be set for them because they were related to toxicity for at least some endpoints. Chemicals screened out for individual endpoints include:

- ***Hyalella azteca* 10-day mortality –** beryllium, butyl benzyl phthalate, chromium, copper, dibutyltin, dimethyl phthalate, di-n-butyl phthalate, mercury, monobutyltin, total chlordanes, total DDTs, tributyltin

- ***Chironomus dilutus* 10-day mortality –**ammonia, antimony, beryllium, bis(2-ethylhexyl) phthalate, dimethyl phthalate, nickel, pentachlorophenol, selenium, vanadium, zinc

- ***Chironomus dilutus* 10-day growth** – ammonia, antimony, beryllium, bis(2-ethylhexyl) phthalate, butyl benzyl phthalate, cadmium, di-n-octyl phthalate, selenium, silver, zinc

- ***Hyalella azteca* 28-day mortality** – 4-methylphenol, antimony, arsenic, beryllium, bis(2-ethylhexyl) phthalate, butyl benzyl phthalate, chromium, copper, dibutyltin, Endrin, lead, monobutyltin, nickel, pentachlorophenol, selenium, tetrabutyltin, tributyltin, vanadium, zinc

- ***Hyalella azteca* 28-day growth** – antimony, arsenic, cadmium, lead, mercury, nickel, selenium, total PAHs

Modeling Results

Finally, the modeling results identified several analytes whose SQV values were greater than the highest concentrations measured for all tests and endpoints. These analytes include butyl benzyl phthalate, dimethyl phthalate, and total chlordanes. No SQVs will be set for these analytes, but site managers can assume that concentrations within the range in this data set are not of concern for benthic organisms.

Table B-3 summarizes all of the analytes that were screened out, the reason for doing so, and the maximum concentration below which site managers can assume that these analytes are not of concern to benthic organisms (where known and applicable).

Table B-2. ANOVA Screening[a]

Analyte	CHR10M SQS/SL1	CHR10M CSL/SL2	CHR10G SQS/SL1	CHR10G CSL/SL2	HYA10M SQS/SL1	HYA10M CSL/SL2	HYA28M SQS/SL1	HYA28M CSL/SL2	HYA28G SQS/SL1	HYA28G CSL/SL2
4-Methylphenol	1**	1**	1**	1**	1**	1**	0	0*		
Aldrin	0	0	0	0			0	0		
alpha-Hexachlorocyclohexane	1	1	0*	1			1	1*		
Ammonia	0*	0*	0	0*	1*	0*	1**	1*		
Antimony	0	0	0	0	1	1*	0	0	0	0*
Arsenic	1**	1**	1**	1**	1*	1**	0	0	0	0
Benzoic acid	1*	1**	1*	1*	1	1				
Beryllium	0	0	0	0	0	0	0	0	1	1
beta-Hexachlorocyclohexane	1**	1**	1	1**			1**	1**		
Bis(2-ethylhexyl) phthalate	0	0	0	0	1	1*	0	0		
Butyl benzyl phthalate	1	1	0	0	0	0	0	0		
Cadmium	1*	0*	0	0	0	1	1	1	0	0*
Carbazole	1**	1**	1**	1**	1	1*	1**	1**		
Chromium	1	1	1	1	0	0*	0*	0*	1**	1**
Copper	1	1**	1	1*	0	0	0	0	0	1*
delta-Hexachlorocyclohexane	1	1*	0*	1			1*	1**		
Dibenzofuran	1**	1**	1*	1**	1	1	1**	1**		
Dibutyltin	1	1*	0*	1	0	0	0	0		
Dieldrin	1	1**	0*	1**			1	1*		
Dimethyl phthalate	0	0*	1	0*	0	0				
Di-n-butyl phthalate	1**	1**	1**	1**	0	0	1	1**		
Di-n-octyl phthalate	1	0	0	0	0	1				
Dioxins/Furans	0	0	0	0			0	0		
Endrin	1	0	1	1*			0	0		
Endrin ketone	1*	1**	0*	1*			1*	1**		
gamma-Hexachlorocyclohexane	0	0	0	0			0	0		
Hexachlorobenzene	0	0	0	0			0	0		
Hexachloroethane	0	0	0	0			0	0		
Lead	1**	1*	1	1*	1	1	0	0	0	0

Analyte	CHR10M SQS/SL1	CHR10M CSL/SL2	CHR10G SQS/SL1	CHR10G CSL/SL2	HYA10M SQS/SL1	HYA10M CSL/SL2	HYA28M SQS/SL1	HYA28M CSL/SL2	HYA28G SQS/SL1	HYA28G CSL/SL2
Mercury	1*	1**	1	1*	0	0	0*	1	0	0
Methoxychlor	0	0	0	0			0	0		
Monobutyltin	1*	1**	1	1**	0	0	0	0		
Nickel	0	0	1*	0*	1	1*	0	0	0	0
Pentachlorophenol	0	0	1**	0	0*	1	0	0		
Phenol	1**	1**	1**	1**	1	1	1**	1**		
Retene					0	0				
Selenium	0	0	0	0			0	0	0	0
Silver	1	1	0	0*	1	0*	1**	1**		
Sulfide	1**	1	1*	1*	1	1*	1	1**		
Tetrabutyltin	1**	1**	1	1**			0	0		
Total Aroclors	1*	1**	1	1*	1	1	1	1**		
Total Chlordanes	1	1**	1	1*	0	0	1*	1**		
Total DDDs	1**	1**	1**	1**	1*	1**	1**	1**		
Total DDEs	1**	1**	1*	1**	1**	0	1**	1**	0*	1
Total DDTs	1	1*	0	1	0	0	1	1*		
Total Endosulfans	0	0	0	0			0	0		
Total PAHs	1**	1**	1**	1**	1	1*	1**	1**	0	0
TPH-Diesel	1**	1**	1**	1**	1**	0	1**	1**		
TPH-Residual	1**	1**	1**	1**	1	0	1**	1**		
Tributyltin	1*	1**	1	1**	0	0	0	0		
Vanadium	0*	0*	1	1			0	0	1	0
Zinc	0	0	0	0	1	1	0	0	0	1*

SQS/SL1 = Sediment Quality Standard/Screening Level 1, CSL/SL2 = Cleanup Screening Level/Screening Level 2
CH10G = *Chironomus* 10-day growth, CH10M = *Chironomus* 10-day mortality,
HY10M = *Hyalella* 10-day mortality, HY28G = *Hyalella* 28-day growth, HY28M = *Hyalella* 28-day mortality
[a] ANOVA results for the relationship between chemical concentration and toxicity for the indicated test and effects level:
0 = not significant, 0* = significant at $p < 0.1$, 1 = significant at $p < 0.05$, 1* = significant at $p < 0.005$, 1** = significant at $p < 0.0005$
A significance level of $p < 0.05$ was used for screening for SQV development.

Table B-3. Summary of Screened Analytes

Chemical Analyte	Reason for Screening	Maximum safe concentration for benthic organisms[a]
1-Methylnaphthalene	Included in Total PAHs	N/A
1,2,3,4-Tetrahydronaphthalene	Infrequently detected	Unknown
1,2,3-Trichloropropane	Infrequently detected	Unknown
1,2,4-Trichlorobenzene	Infrequently detected	Unknown
1,2-Dichlorobenzene	Infrequently detected	Unknown
1,2-Dichloroethane	Infrequently detected	Unknown
1,4-Dichlorobenzene	Infrequently detected	Unknown
2-Methylnaphthalene	Included in Total PAHs	N/A
2,3,4,5-Tetrachlorophenol	Infrequently detected	Unknown
2,3,4,6-Tetrachlorophenol	Infrequently detected	Unknown
2,4,5-Trichlorophenol	Infrequently detected	Unknown
2,4-D	Infrequently detected	Unknown
2,4-DB	Infrequently detected	Unknown
2,4-Dichlorophenol	Infrequently detected	Unknown
2,4-Dimethylphenol	Infrequently detected	Unknown
2,4-Dinitrotoluene	Infrequently detected	Unknown
2-Chloronaphthalene	Infrequently detected	Unknown
2-Chlorophenol	Infrequently detected	Unknown
2-Methylphenol	Infrequently detected	Unknown
4-Chloro-3-methylphenol	Infrequently detected	Unknown
4-Nitroaniline	Infrequently detected	Unknown
4-Stigmasten-3-one	Infrequently detected	Unknown
7,10,13-Hexadecatrienoicacid	Infrequently detected	Unknown
9-Hexadecenoicacid	Infrequently detected	Unknown
Abietic acid	Infrequently detected	Unknown
Acenaphthene	Included in Total PAHs	N/A
Acenaphthylene	Included in Total PAHs	N/A
Acid volatile sulfides	Derived parameter	N/A
Aldrin	No relationship to toxicity	Up to a maximum concentration of 690 µg/kg
alpha-Chlordane	Included in Total chlordanes	N/A
alpha-Endosulfan	Included in Total endosulfans	N/A

alpha-Hexachlorocyclohexane	Not enough hits	Minimal data suggests possible toxicity over 5 µg/kg
Aluminum	Crustal element	N/A
Aniline	Infrequently detected	Unknown
Anthracene	Included in Total PAHs	N/A
Aroclors (all)	Included in Total PCBs	N/A
Benz(a)anthracene	Included in Total PAHs	N/A
Benzene	Infrequently detected	Unknown
Benzo(a)pyrene	Included in Total PAHs	N/A
Benzo(b)fluoranthene	Included in Total PAHs	N/A
Benzo(ghi)perylene	Included in Total PAHs	N/A
Benzo(k)fluoranthene	Included in Total PAHs	N/A
Benzyl alcohol	Infrequently detected	Unknown
Beryllium	Not enough hits	Minimal data shows no evidence of toxicity up to 1.5 mg/kg (maximum concentration detected)
beta-Endosulfan	Included in Total endosulfans	N/A
Bis(2-chloroethyl) ether	Infrequently detected	Unknown
Butyl benzyl phthalate	Modeling identified no toxicity	Up to a maximum concentration of 2800 µg/kg
Calcium	Crustal element	N/A
Caprolactam	Infrequently detected	Unknown
Carbon disulfide	Infrequently detected	Unknown
Chlordane	Included in Total chlordanes	N/A
Chlorobenzene	Infrequently detected	Unknown
Chloroform	Infrequently detected	Unknown
Chloromethane	Infrequently detected	Unknown
Chrysene	Included in Total PAHs	N/A
cis-1,2-Dichloroethene	Infrequently detected	Unknown
cis-Chlordane	Included in Total chlordanes	N/A
cis-Nonachlor	Included in Total chlordanes	N/A
Dehydroabietic acid	Infrequently detected	Unknown
delta-Hexachlorocyclohexane	Not enough hits	Minimal data suggests possible toxicity over 2.4 µg/kg
Dibenz(ah)anthracene	Included in Total PAHs	N/A
Dichloromethane	Infrequently detected	Unknown
Diethyl phthalate	Infrequently detected	Unknown

Dimethyl phthalate	Modeling identified no toxicity	Up to a maximum concentration of 580 µg/kg
Dioxins/furans	No relationship to toxicity	Up to a maximum concentration of 28,000 ng/kg
Endosulfan sulfate	Included in Total endosulfans	N/A
Endrin	Not enough hits	Minimal data shows no clear toxicity up to 40 µg/kg (maximum detected value)
Endrin aldehyde	Infrequently detected	Unknown
Ethylbenzene	Infrequently detected	Unknown
Fluoranthene	Included in Total PAHs	N/A
Fluorene	Included in Total PAHs	N/A
gamma-Chlordane	Included in Total chlordanes	N/A
gamma-Hexachlorocyclohexane	Not enough hits	Minimal data shows no clear toxicity up to 11 µg/kg (maximum detected value)
gamma-Sitosterol	Infrequently detected	Unknown
Grain size	Physical parameter	N/A
Heptachlor	Included in Total chlordanes	N/A
Heptachlor epoxide	Included in Total chlordanes	N/A
Heptachlorodibenzofurans	Included in Dioxins/furans	N/A
Heptachlorodibenzo-p-dioxins	Included in Total dioxins/furans	N/A
Hexachlorobutadiene	Infrequently detected	Unknown
Hexachlorobenzene	No relationship to toxicity	Up to a maximum concentration of 260 µg/kg
Hexachlorodibenzofurans	Included in Dioxins/furans	N/A
Hexachlorodibenzo-p-dioxins	Included in Dioxins/furans	N/A
Hexachloroethane	No relationship to toxicity	Up to a maximum concentration of 1500 µg/kg
Indeno(123-cd)pyrene	Included in Total PAHs	N/A
Iron	Crustal element	N/A
Isophorone	Infrequently detected	Unknown
Isopimaric acid	Infrequently detected	Unknown
m,p-Xylene	Infrequently detected	Unknown
Magnesium	Crustal element	N/A
Manganese	Crustal element	N/A
MCPA	Infrequently detected	Unknown
MCPP	Infrequently detected	Unknown
Methoxychlor	No relationship to toxicity	Up to a maximum concentration of 34 µg/kg
Methyl iodide	Infrequently detected	Unknown

Methyl tert-butyl ether	Infrequently detected	Unknown
Methylethyl ketone	Infrequently detected	Unknown
Mirex	Infrequently detected	Unknown
N-Nitrosodiphenylamine	Infrequently detected	Unknown
Naphthalene	Included in Total PAHs	N/A
o-Xylene	Infrequently detected	Unknown
o,p'-DDD	Included in Total DDDs	N/A
o,p'-DDE	Included in Total DDEs	N/A
o,p'-DDT	Included in Total DDTs	N/A
Octachlorodibenzofuran	Included in Dioxins/furans	N/A
Octachlorodibenzo-p-dioxin	Included in Dioxins/furans	N/A
Oxychlordane	Included in Total chlordanes	N/A
p,p'-DDD	Included in Total DDDs	N/A
p,p'-DDE	Included in Total DDEs	N/A
p,p'-DDT	Included in Total DDTs	N/A
Pentachlorodibenzofurans	Included in Dioxins/furans	N/A
Pentachlorodibenzo-p-dioxins	Included in Dioxins/furans	N/A
Perylene	Infrequently detected	Unknown
Phenanthrene	Included in Total PAHs	N/A
Phytol	Infrequently detected	Unknown
Pimaric acid	Infrequently detected	Unknown
Potassium	Crustal element	N/A
Pristane	Infrequently detected	Unknown
Pyrene	Included in Total PAHs	N/A
Retene	No relationship to toxicity	Up to a maximum concentration of 810,000 µg/kg
Sandaracopimaric Acid	Infrequently detected	Unknown
Sodium	Crustal element	N/A
Styrene	Infrequently detected	Unknown
TEQs (dioxin/furan/PCBs)	Derived parameter not applicable to benthos	N/A
Tetrachlorodibenzofurans	Included in Dioxins/furans	N/A
Tetrachlorodibenzo-p-dioxins	Included in Dioxins/furans	N/A
Thallium	Infrequently detected	Unknown

Toluene	Infrequently detected	Unknown
Total benzofluoranthenes (b+j+k)	Included in Total PAHs	N/A
Total chlordanes	Modeling identified no toxicity	Up to a maximum concentration of 670 µg/kg
Total endosulfans	No relationship to toxicity	Up to a maximum concentration of 240 µg/kg
Total organic carbon	Natural material	N/A
Total solids	Physical parameter	N/A
trans-Chlordane	Included in Total chlordanes	N/A
trans-Nonachlor	Included in Total chlordanes	N/A
Trichloroethene	Infrequently detected	Unknown
Vanadium	Not enough hits	Minimal data shows no evidence of toxicity up to 41 mg/kg (maximum concentration measured)
Xylenes	Infrequently detected	Unknown

[a] Does not address potential bioaccumulation toxicity to wildlife, fish, or humans.