DEPARTMENT OF
ECOLOGY
State of Washington

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

November 2011
Publication No. 11-09-054

Publication and Contact Information

This report is available on the Department of Ecology's website at
www.ecy.wa.gov/biblio/1109054.html

For more information contact:

Toxics Cleanup Program
P.O. Box 47600
Olympia, WA 98504-7600

Phone: 360-407-7170

Washington State Department of Ecology - www.ecy.wa.gov

- Headquarters, Olympia 360-407-6000
- Northwest Regional Office, Bellevue 425-649-7000
- Southwest Regional Office, Olympia 360-407-6300
- Central Regional Office, Yakima 509-575-2490
- Eastern Regional Office, Spokane 509-329-3400

If you need this document in a format for the visually impaired, call the Toxics Cleanup Program at 360-407-7170. Persons with hearing loss can call 711 for Washington Relay Service. Persons with a speech disability can call 877-833-6341

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

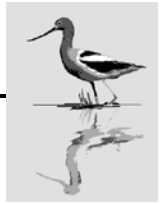
November, 2011

Prepared for:

Washington Department of Ecology
Olympia, WA

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

Avocet Consulting



Under contract to:
Ecology & Environment
and Hart Crowser
Seattle, WA

Acknowledgements

This report was prepared by Dr. Teresa Michelsen, under contract to E&E and Hart Crowser and the Washington Department of Ecology, on behalf of the RSET agencies. From 2007–2009, 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 Ecology sediment staff, including Jim Reese (NW Region Corps), Stephanie Stirling (Seattle District Corps), Dave Bradley, Chance Asher, Russ McMillan, and Laura Inouye (WA Dept. of Ecology). Funding was provided by the Washington Department of Ecology, Oregon Department of Environmental Quality, and NW Region Corps of Engineers.

Table of Contents

Tables

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 values (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 SQVs with greater reliability than existing SQV sets using the Floating Percentile Model (FPM).

As part of this initial effort, it was determined that freshwater apparent effects thresholds (AETs) were not as reliable as the marine AETs; specifically, they were less conservative. Marine systems are chemically buffered and are far more similar to one another than freshwater areas of the state, which have a wide range of chemical, geological, and habitat types. This similarity between marine areas lends itself well to the mathematical methods used to calculate the AETs. However, because of the variation among freshwater areas, selection of the highest no-hit value as the AET allowed an unacceptable degree of toxicity. Therefore, a different mathematical approach was used for calculating the SQVs that would ensure appropriately low levels of toxicity.

As a result, there are some notable differences between the marine and freshwater SQVs:

- Because the mathematical models used to calculate the SQVs are different, the values cannot be directly compared. For example, the AETs are calculated on a single-chemical basis, while the FPM values are calculated on a multivariate basis, looking at all chemicals together.

- In the 20 years since the marine AETs were first calculated, it has been determined that organic-carbon normalization does not improve the reliability of the SQVs. This was confirmed again in 2003 during the development of proposed SQVs for Ecology (SAIC and Avocet 2003). Therefore, the proposed freshwater SQVs are calculated on a dry weight basis.

- Due to differences in the larger geographic range encompassed by the freshwater SQVs, differences in sources (industries and chemicals) in marine vs. freshwater areas of the state, and differences in bioavailability and toxicity of certain chemicals (especially metals) in freshwater vs. marine systems, there are different chemicals included on each list and different levels for the same chemicals. These differences are based on actual field conditions and are to be expected.

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 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 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) governing cleanup of sediment sites in Washington State.

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

Nearly all of these goals were achieved during the update process. 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 three acute and two 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. While the data are representative of the majority of freshwater sediment sites encountered in the northwest, it is recognized that benthic toxicity at sites with unique geochemical characteristics will differ and the SQVs are not representative of those sites (e.g., bogs, alpine wetlands, sites with mining, milling or smelting activities, substantial waste deposits, or with unique pH, alkalinity, or other geochemical characteristics). Freshwater bioassays should be used to assess toxicity under these conditions.

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 correctly predicted overall 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 compared to other available SQV sets, greatly improving the implementability and cost-effectiveness of the SQVs. In addition, at the higher effects levels, the FPM values are also able to detect more of the 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 for public review, incorporation into the SEF, and MTCA/SMS rule revision (Table ES-1). The method used to develop these values is based on specific assumptions about the levels of risk and error that are considered acceptable at each effects level, and provides the opportunity for revision of the SQVs if alternative policy choices are made during the public review process.

These values were developed to protect only against toxicity to the benthic community in freshwater environments. They are not protective of bioaccumulative effects to humans, wildlife, or fish.

Table ES-1. Proposed Sediment Quality Values

Analyte	SQS/SL1[a]	CSL/SL2[b]
Conventional Pollutants (mg/kg)		
Ammonia	230	300
Total sulfides	39	61
Metals (mg/kg)		
Arsenic	14	120
Cadmium	2.1	5.4
Chromium	72	88
Copper	400	1200
Lead	360	> 1300
Mercury	0.66	0.8
Nickel	26	110
Selenium	11	> 20
Silver	0.57	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	900	1100
Dibenzofuran	200	680
Dibutyltin	910	130000
Dieldrin	4.9	9.3
Di-n-butyl phthalate	380	1000
Di-n-octyl phthalate	39	> 1100
Endrin ketone	8.5	**
Monobutyltin	540	> 4800
Pentachlorophenol	1200	> 1200
Phenol	120	210
Tetrabutyltin	97	> 97
Total DDDs	310	860
Total DDEs	21	33
Total DDTs	100	8100
Total PAHs	17000	30000
Total PCB Aroclors	110	2500
Tributyltin	47	320
Bulk Petroleum Hydrocarbons (mg/kg)		
TPH-Diesel	340	510
TPH-Residual	3600	4400

[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 SQS/SL1 concentration.

This page was left blank intentionally

1. Introduction

This report presents the results of the 2010 recalculation of freshwater sediment quality guidelines (SQVs) for Washington, Oregon, and Idaho. The SQVs update was begun 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 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 and completion of these SQVs for use in cleaning up contaminated sediment sites under the Sediment Management Standards (SMS) and Model Toxics Control Act (MTCA).

1.1 Freshwater SQV Early Development (2002–2003)

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 all 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 toxicity.
- A reliability analysis of eight existing North American 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 neither existing freshwater SQV sets nor the marine AETs were able to correctly predict both toxic and non-toxic samples with an acceptable degree of reliability in freshwater environments, 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 freshwater SQVs based on a newly developed 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 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 Update of the Freshwater SQVs (2007–2011)

In 2007, RSET undertook an update of Ecology's freshwater SQVs for inclusion in the SEF, beginning a four-year process that concluded in this report. 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 produced SQVs with improved reliability and a complete set of acute and chronic endpoints with reliable SQVs. Ecology conducted further review by the MTCA/SMS Science Panel and a national scientific peer review in 2009–2010, and EPA Region 10 also provided statistical input. The results of all of these efforts are reflected in this report.

1.3 Public Outreach and Peer Review

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
- 2011 Advanced Sediment Cleanup Conference, Seattle, WA

In addition, Ecology's rule advisory groups (Sediment Workgroup and MTCA/SMS Advisory Group) for the MTCA/SMS rule revisions reviewed the method in a series of meetings in 2010, the MTCA/SMS Science Panel reviewed the approach in 2010 and 2011, and Ecology requested a review of the method and draft report from four national-level scientific peer reviewers. Additional formal public review and comment will occur during the public review period associated with the SMS rule revision.

1.4 Supplemental Electronic Files

A variety of additional electronic files are available on Ecology's website providing the underlying data set, modeling spreadsheets, and statistical evaluations summarized in Sections 2–4 of this report:

- **Station Locations** – A complete list of the stations included in the data set (Figure 2-1) and their latitudes/longitudes can be found in the spreadsheet "LatLongs.xls".

- **Final Chemistry Data Sets –** The complete chemistry data set summarized in Table 2-4 can be found in the spreadsheet "Final Chemistry.xls". Individual data sets for each bioassay endpoint can be found in spreadsheets of the same name appended with the bioassay abbreviations, e.g., "Final Chemistry-CH10G.xls".

- **Toxicity Test Results –** Results of the toxicity tests in the form of hit (1) or no-hit (0) designations for each sample summarized in Table 2-3 can be found in the spreadsheet "BioHitNoHit.xls". Hit/no-hit files for each of the individual bioassay endpoints can be found in spreadsheets of the same name appended with the bioassay abbreviations, e.g., "BioHitNoHit-CH10G.xls".

- **FPM Step 1. Initial Data Processing** – The results of the first FPM model spreadsheet, which screens, sums, summarizes, and formats the chemistry data for modeling, can be found in the spreadsheet "FPMData.xls". Results of this step for each of the individual bioassay endpoints can be found in spreadsheets of the same name appended with the bioassay abbreviations, e.g., "FPMData-CH10G.xls". One additional spreadsheet, "FPMDataGroups.xls", is also included showing how chemical classes were summed for modeling. The output table of the FPMCalc spreadsheet is imported into the second modeling spreadsheet described below.

- **FPM Step 2. ANOVA Screening –** The results of the second FPM model spreadsheet, which evaluates the association of each chemical with toxicity in the data set, can be found in the spreadsheets named "FPMAnova*.xls". There is one of these spreadsheets for each bioassay endpoint and each effects endpoint, e.g., "FPMAnova-CH10G-SL1.xls". This spreadsheet includes a summary table showing the strength of each chemical's association with toxicity in the data set, the ability to select or deselect chemicals for continued modeling based on these results, and a set of worksheet tabs showing the hit and no-hit distributions for each chemical on which the analysis is based. The output table of the FPMAnova spreadsheet is imported into the third modeling spreadsheet described below.

- **FPM Step 3. Model Calculations** – The results of the third FPM model spreadsheet, which calculates the SQVs (summarized in Section 3.3) and evaluates their predictive reliability (summarized in Section 4.1), can be found in the spreadsheets named "FPMCalc*.xls". There is one of these spreadsheets for each bioassay endpoint and each effects endpoint, e.g., "FPMCalc-CH10G-SL1.xls". As explained in Section 3.1, the model can be run in two different ways, and the results of both are provided on the "Data Storage" tab. The row ultimately selected as the basis of the SQVs presented in this report is highlighted on that tab.

- **Supplemental Statistics** – Spreadsheets for calculating the supplemental statistical evaluations discussed in Section 4.3 can be found in the spreadsheets named "SuppStatistics*.xls". There is one of these spreadsheets for each bioassay endpoint and each effects endpoint, e.g., "SuppStatistics-CH10G-SL1.xls", as well as one for the complete draft SQS/SL1 and one for the CSL/SL2 SQV sets. The template spreadsheet was provided by EPA Region 10, and includes a wide variety of additional statistical measures not used in this report. However, they may be of interest to readers.

2. Database Development

The following sections describe the collection, screening, processing, and assembly of the data set used in the FPM model runs. The resulting data set is also summarized. Additional electronic files containing station locations and the underlying bioassay and chemistry data sets are also available, described in Section 1.4.

2.1 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 (see Appendix B). 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 could not 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 (growth and mortality endpoints). 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.

2.2 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.

Synoptic Samples – Data were only used if chemistry analyses and bioassays were run on splits from the same homogenized sample. Surveys were not included if chemistry and bioassay samples were collected at different times, from different locations, or from different grab samples.

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. At least semivolatiles (e.g., Method 8270) and a complete set of metals was selected as a minimum guideline for including a survey or station, consistent with other national criteria development efforts. Metals and semivolatiles both are significantly associated with 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. For some surveys, different stations had varying analyte lists. In these surveys, only those stations with adequate analyte lists were retained. Eleven surveys and an additional 9 stations from one survey were screened out due to insufficient analyte lists. Unfortunately, many eastern Washington surveys fell into this category, having only conventionals and/or a few metals (see Appendix B) colocated with bioassay data.

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. We were unable to obtain data for 5 small surveys (<10 samples each).

Minimum amount of data **-** For development of SQVs, a minimum number of data points is required. A minimum of 30 detected values was chosen as the lower limit for inclusion on the analyte list at the initiation of the project. Depending on the chemical distributions and range of bioassay responses in the data set, a larger number (up to 100) may be required for some projects; however, this value was chosen to be as inclusive as possible. Several of these chemicals were later removed from the dataset when it was determined that there were only a few toxic stations among the 30+ detected values for that chemical, not enough to develop a reliable criterion.

Chemicals with <30 detected data are listed in Appendix B. These 61 chemicals included primarily volatile or unusual compounds not generally expected to be found for most projects, as well as some herbicides/pesticides not widely used in the Pacific Northwest. However, should they be important for a specific site, bioassay testing is recommended for evaluation of their potential toxicity.

Nontoxicity **-** 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. While some of these compounds can affect the toxicity of other chemicals at certain sites and can be useful in risk assessments, they are not themselves toxic and thus do not require the development of SQVs. Certain conventional analytes, such as grain size parameters and acid-volatile sulfides, were screened out because they are physical parameters or derived quantities. Other derived quantities frequently present in data sets, such as dioxin toxicity equivalency quotients (TEQs) for human health, were also not included, because they are not related to benthic toxicity. These analytes are listed in Appendix B.

Chemistry quality assurance – All chemistry data were qualified to "QA2" level, as defined in Ecology (1989), a high level of quality assurance designed to support rule-making or litigation purposes. Quality assurance was conducted consistent with the SEF (2009) and in accordance with PSEP QA2 (PTI 1989), DMMP (2009), and US EPA (1986, 1987a,b,c, 1999, 2004, 2007) manuals. 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 – All bioassay data were subjected to a QA2 level of review using an in-house checklist and verification of all original laboratory data and calculations. The review included:

- General project and test endpoint information
- Chain of custody, holding times, and holding conditions
- Sources of organisms and species
- Number of replicates
- Whether all aspects of the protocols were followed/non-standard protocol elements
- Whether all required water quality parameters were measured and within control limits
- Positive control toxicant, control charts, and whether the LC50 was within control limits
- Source of the negative control and whether it was within control limits
- Whether reference samples were within control limits
- Hand-check of all calculations

Six surveys, comprising 46 stations, did not meet one or more minimum QA requirements. Many of these surveys also had an insufficient analyte list as described above (see Appendix B).

2.3 Normalization and Summing

Organic carbon normalization - 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 inappropriate in some situations with large quantities of anthropogenically derived organic carbon or under natural conditions with very low amounts of organic carbon. Consistent with regional dredging guidelines and all other national SQVs, the current SQVs are calculated on a dry weight normalized basis.

Petroleum hydrocarbons **-** 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 far more strongly associated with petroleum toxicity than PAHs, although there were no TPH data for many stations (see Appendix D for details of the model runs). Therefore, both were retained in the model runs and the two together provided better reliability than either one alone.

Chemical Classes **-** 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 of the sums, which were not included as individual chemicals in the model runs to reduce covariance among variables.

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 method detection limit and summed with the other constituents.

- Unusually high detection limits (e.g., due to interference noted in QA/QC reports) 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 as part of summed classes were evaluated by the workgroup, including eliminating 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 other available statistical methods, which each have other limitations and would potentially need to be applied differently depending on the distribution of and/or number of nondetects in each individual chemical sum.

2.4 Comparison to Control vs. Reference

In the marine sediment cleanup and dredging programs, bioassay controls are used to evaluate the performance of the test, and bioassay test samples are compared to reference sediment samples from clean areas of Puget Sound. The reference samples are intended to "correct" for effects that physical parameters of the sediment may have on the test animal. However, reference areas have not been identified in freshwater areas of the state despite significant efforts by the agencies, in part due to much greater variability of freshwater environments and in part due to the lack of uncontaminated upstream areas.

Based on the results of SAIC and Avocet (2002) as well as updated evaluations conducted with the current data set (see Section 3.2 and Appendix D), 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, depending on the endpoint, approximately two-thirds of the 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. Appendix D, section D3 covers this issue in more detail.

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 in the future (Stirling and RSET 2008), 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.

2.5 Bioassay Tests and Endpoints

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

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

In Washington State sediment programs, identification of adverse biological effects 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. Development of the thresholds for each bioassay endpoint is presented in Appendix C. Data transformations, selection of null hypotheses, and appropriate statistical tests (depending on the data distribution) are identical to those currently in use by RSET for marine sediment data (Michelsen and Shaw 1996, Fox et al. 1998). In all cases, "statistically significant" means a statistical difference from a control sample at an alpha level of 0.05.

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%[a]	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)

2.6 ANOVA Analyte Screening

Once the individual biological tests and endpoints had been selected, 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 $\leq$ 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 using ANOVA 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. These chemicals could be retained in the model, but it would run more slowly and give the same results.

Detailed results of the ANOVA screening evaluations, which were conducted separately for each chemical, effects level, and endpoint combination, are provided 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 apparent 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 associated with 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.

2.7 Final Data Set

Figure 2-1 shows the station locations included in the final data set, identifying hit and no-hit stations. The data set comprises 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). Most of the stations are located in three general areas: freshwater locations near Seattle, WA and Portland OR, and the upper Columbia and Spokane Rivers. There are also a number of stations downstream of the Willamette River in the Columbia River. With the

exception of the lower Columbia River, which is mainly no-hit stations, hit stations are fairly evenly distributed throughout the data set in these regions. Appendix A provides a list of surveys included in the final data set, including the state and region, number of stations for each bioassay, analyte classes included in the survey, and references.

The numbers of stations for each bioassay endpoint are shown in Table 2-3 (samples that failed quality assurance evaluation are not included). Table 2-3 also shows the number and percentage of stations associated with biological hits for each bioassay and effects level. Overall, toxicity was observed at 12–33% of the stations at the lower SQS/SL1 level and at 7–15% of the stations at the higher CSL/SL2 level.

Table 2-3. Bioassays and Endpoints in Final Data Set

Test	No. of Samples	SQS/SL1[a]	CSL/SL2[a]
Hyalella azteca			
10-day mortality	366	89 (24%)	52 (14%)
Hyalella azteca			
28-day mortality	312	47 (15%)	27 (7%)
Hyalella azteca			
28-day growth	79	26 (33%)	12 (15%)
Chironomus dilutus			
10-day mortality	568	85 (15%)	41 (7%)
Chironomus dilutus			
10-day growth	525	65 (12%)	49 (9%)

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

Table 2-4 provides a summary of the concentration distributions for each of the chemicals detected more than 30 times in the data set, including chemicals screened out as described above. For chemicals detected less than 30 times, see Appendix B. In each case, the median was less than the mean, usually by a substantial amount. This pattern indicates a right-skewed data set as would be expected for an environmental data set containing highly contaminated areas. For most chemicals (particularly those remaining after the screening described above), the concentration ranges were quite large, indicating inclusion of both clean and contaminated areas.

Figure 2-1. Station Locations

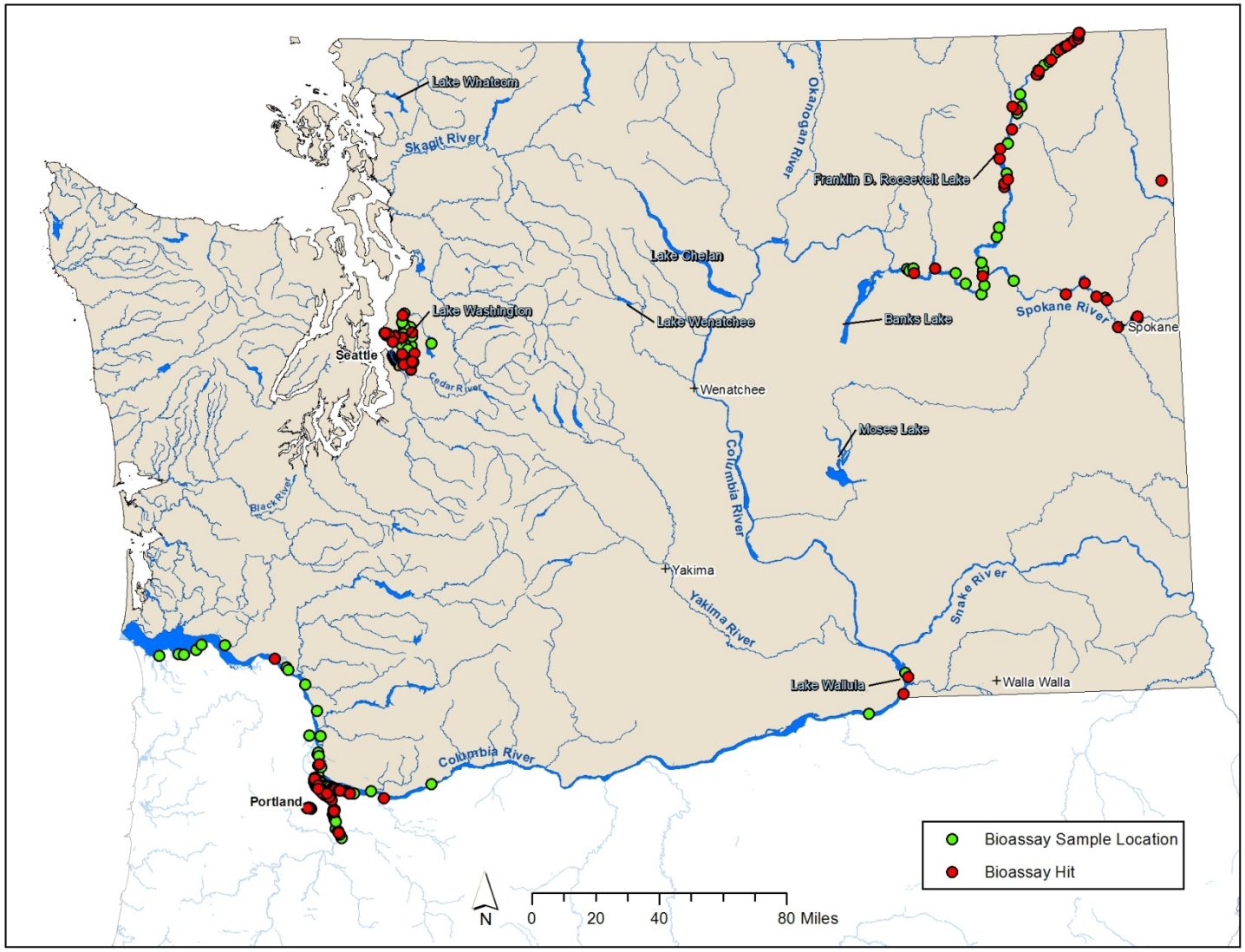


Table 2-4. Chemical Distributions[a]

Analyte	N	Minimum	Median	Mean	Maximum
Conventional Pollutants (mg/kg)					
Ammonia	424	0.050	69	87	780
Total sulfides	329	0.20	7.1	67	7700
Metals (mg/kg)					
Antimony	342	0.050	0.20	3.1	310
Arsenic	613	0.48	4.4	11	1200
Cadmium	528	0.040	0.34	0.97	40
Chromium	533	3.8	30	35	350
Copper	559	3.3	39	120	11000
Lead	519	0.62	26	86	1400
Mercury	535	0.006	0.085	0.29	43
Nickel	544	5.0	23	27	590
Selenium	233	0.040	0.14	0.91	20
Silver	409	0.024	0.21	0.39	4.5
Zinc	568	15	120	390	14000
Organic Chemicals (µg/kg)					
4-Methylphenol	151	4.0	28	200	6300
Aldrin	77	0.052	0.86	14	690
alpha-Hexachlorocyclohexane	66	0.047	0.26	0.83	10
Benzoic acid	64	20	300	810	4200
beta-Hexachlorocyclohexane	131	0.16	1.6	3.0	26
bis(2-Ethylhexyl)phthalate	303	4.2	260	2800	440000
Butylbenzyl phthalate	172	2.7	44	140	2800
Carbazole	218	2.1	25	5000	480000
delta-Hexachlorocyclohexane	48	0.092	0.36	1.1	21
Dibenzofuran	356	0.20	11	8300	2200000
Dibutyltin	124	0.017	20	2600	160000
Dieldrin	61	0.079	0.42	7.9	360
Dimethyl phthalate	47	4.5	49	98	580
Di-n-butyl phthalate	203	4.0	15	92	1800
Di-n-octyl phthalate	62	3.1	40	250	4300
Dioxins/furans (ng/kg)	73	2.4	130	860	28000
Endrin	38	0.043	2.5	7.0	39
Endrin ketone	60	0.078	0.85	2.9	90
gamma-Hexachlorocyclohexane	48	0.20	1.9	2.8	11
Hexachlorobenzene	127	0.26	1.4	4.3	260
Hexachloroethane	44	0.38	1.8	38	1500
Methoxychlor	48	0.048	2.3	4.9	34
Monobutyltin	141	0.16	11	100	4800
Pentachlorophenol	81	0.81	15	290	16000
Phenol	120	3.5	16	47	770
Retene	38	11	1200	39000	810000
Tetrabutyltin	54	0.33	3.0	40	770
Total Chlordanes	218	0.042	1.3	15	670
Total DDDs	318	0.046	4.7	68	3000
Total DDEs	321	0.087	3.0	25	2500

Analyte	N	Minimum	Median	Mean	Maximum
Total DDTs	263	0.077	3.1	130	13000
Total Endosulfans	41	0.048	0.54	8.8	240
Total PAHs	609	0.20	970	120000	36000000
Total PCB Aroclors	320	0.85	72	330	27000
Tributyltin	190	0.029	24	3600	300000
Bulk Petroleum Hydrocarbons (mg/kg)					
TPH-Diesel	184	14	150	870	39000
TPH-Residual	206	16	490	1200	18000

[a] Detected values only, prior to chemical screening described above.

This page was left blank intentionally

3. SQV Calculations

The basic concept behind the FPM is to select an optimal percentile of the data set that provides a specified 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 false negative rate (the false negative rate is not allowed to increase).

Once each chemical has been individually adjusted upward to the point where it begins to show an association with toxicity, the false positives will have been significantly reduced while retaining the same false negative rate. In this manner, SQVs can be developed for a number of different target false negative rates (e.g., 0–30%), allowing the trade-offs between false negatives and false positives to be evaluated and a final set of SQVs to be selected. The model spreadsheets for each bioassay endpoint and effects level are available as supplemental electronic files, as described in Section 1.4. Each spreadsheet contains instructions for running the model and the original data set used, to allow duplication of the results.

3.1 Modeling Approach

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.

The first three bullets above are conducted in preparation for running the model, and are described in Section 2. The model carries out the final three bullets within the spreadsheets.

Excel Spreadsheets. Calculation of SQVs occurs through an iterative automated process using Excel Visual Basic macros, as follows:

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 process.

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 also 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 a point above which false negatives increase or a level at which it has no more false positives.

The model can be run in two manners: 1) for a single selected false negative rate (e.g., 20%), or 2) for a range of false negative rates with a given interval (e.g., 0–30% with steps of 5%). If a range is chosen, the model repeats all of the steps above and creates a new row for each false negative rate in the range (e.g., 0, 5, 10, 15, 20, 25, and 30%). When the model is run for a range of false negative rates, it goes through an additional process after calculating all the rows, as follows:

9. Find the lowest value for each chemical among all the rows and restart the calculations using this set of lowest values. Follow steps 1–8 until the lowest false negative rate target is reached.

10. Start the next row using the results of the first row. Follow steps 1–8 until that row's false negative target has been reached. Repeat for all of the false negative targets in the range until a new set of rows is generated.

This second pass through the data set helps deal with the effects of covariance. Although the initial model assumes that all variables are independent of one another, in reality, some chemicals will covary or be colocated and affect each others' results. This can cause a "seesaw" effect, where one chemical concentration is low in some rows while the associated chemical's concentration is high, and vice versa in other rows. Steps 9 and 10 help equalize these effects by finding the lowest concentrations for all chemicals, which may reflect the values they would have in the absence of other covarying or colocated chemicals, and working evenly back and forth between the chemicals.

Through this process, it is possible to identify those analytes having the greatest association with toxicity in the data set (those whose concentrations cannot be increased without increasing false negatives), and those chemicals having little or no association with 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).

Hit/No-Hit Definitions. The model was run separately for each individual bioassay endpoint at both the SQS/SL1 and CSL/SL2 effects levels shown in Table 2-2. This allows greater evaluation of the individual bioassay endpoints – for example, which ones behave similarly, which chemical groups each responds to, and which endpoints are most sensitive and reliable.

Pooled endpoints could also be used, which requires assigning one overall hit/no-hit value to a station based on the performance of all the bioassays at that station. For example, a station could be identified as a hit if any one bioassay showed a hit, and there are a number of other decision rules that could also be chosen. However, for development of the SQVs, this approach was not used because of the historical nature of the data set. Stations had varying numbers of bioassays, ranging from 1–5, and many of the stations did not meet current decision rules required by the SMS (at least three bioassays, both acute and chronic). For site-specific evaluations where all stations have the same set of bioassays, a pooled endpoint could effectively be used.

3.2 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, and results of each are included in Appendix D:

- **Petroleum Hydrocarbons.** The model was run using 1) total PAHs, 2) TPH-diesel and TPH-residual, and 3) both combined for two different data sets. The large data set included all data in the database, for which all stations had PAH data but only about 1/3 had TPH data. The small data set included only those stations that had both PAH and TPH data.

- **Regional Differences.** The model was run for the entire data set, as well as separately for 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 was intended to evaluate whether different SQVs would be appropriate for these georegions.

- **Comparison to Control vs. Reference.** The subset of the data set that includes reference data was used to evaluate the reliability of comparison to control vs. comparison to reference, to test the previous finding (SAIC and Avocet, 2003) that comparison to control provides similar or better reliability than comparison to reference, given the current nature of the data set.

- **Blank-Correction.** It was determined during the quality assurance review that the data sets had not all been blank-corrected in the same manner, and that some common laboratory contaminants rarely found in the environment were inappropriately appearing in the SQV tables. This issue was addressed by re-qualifying all of the historic data sets in a consistent manner, using EPA Contract Laboratory Protocols, and then rerunning the model to assess the effects.

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

- **Petroleum Hydrocarbons.** 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. The TPH measures were more reliable; however, TPH data were missing for many data sets, leading to improved reliability when both were included.

- **Regional Differences.** East- and west-side data were combined into a single data set. The reliability of the different regions varied by endpoint and was highly dependent on the amount of data available on the east side. It may be possible in the future to calculate SQVs for different geographic regions once more data are available.

- **Comparison to Control vs. Reference.** Current results for comparison to reference vs. comparison to control were consistent with SAIC and Avocet (2003), indicating that comparison to control was at least as reliable as comparison to reference and allowed use of a much larger data set. Therefore, the model was run based on comparison to control.

- **Blank-Correction.** For stations with detected concentrations in the blanks, revising the qualifiers consistent with the approach specified by the EPA Contract Laboratory Protocols eliminated analytes from the SQV list known to be common laboratory contaminants (e.g., acetone, methylene chloride) that had previously been associated with a significant number of false positives.

3.3 Final Model Results

Tables 3-1 and 3-2 show the resulting FPM values for each endpoint based on the modeling approach described above and the reliability assessment described in Section 4. These values best meet the reliability goals of Ecology and the RSET SQV development workgroup. “Greater than” signs (>) indicate that the toxicity value for that chemical and endpoint is greater than any of the concentrations in the database, and the maximum concentration is shown in the table.

Table 3-1. 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	0.57	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 Chlordanes	> 670	> 670	--	--	> 670
Total DDDs	860	2500	310	--	2500
Total DDEs	910	910	21	> 5.7	910
Total DDTs	> 13000	100	--	--	8100
Total PAHs	30000	45000	17000	--	330000
Total PCB Aroclors	3100	3400	110	--	3400
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 3-2. 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	120	200	14	16
Cadmium	6.3	13	13	> 23	> 23
Chromium	220	220	> 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	0.64	4.1	--	1.7
Zinc	> 14000	--	> 4200	3200	> 14000
Organic Chemicals (µg/kg)					
4-Methylphenol	> 6300	2000	2400	--	260
Benzoic acid	--	2900	3800	--	--
beta-Hexachlorocyclohexane	11	11	--	--	11
bis(2-Ethylhexyl)phthalate	> 440000	--	22000	--	> 440000
Butylbenzyl phthalate	> 2800	> 2800	> 1500	--	> 2800
Carbazole	1400	900	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	>1800	> 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 Chlordanes	24	> 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
Total PCB Aroclors	3400	3400	2500	--	3400
Tributyltin	9300	320	47	--	> 9300
Bulk Petroleum Hydrocarbons (mg/kg)					
TPH-Diesel	510	510	2100	--	1300
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

4. Reliability Assessment

A reliability assessment was conducted following derivation of the SQVs. The assessment was conducted in two parts – first, candidate SQVs were evaluated using standard measures of reliability such as false positives, false negatives, and overall reliability, and these results were used to select the values that appear in Tables 3-1 and 3-2. In addition, these reliability measures were used to compare the FPM SQVs with other freshwater SQV sets available in North America.

Subsequently, EPA and others recommended additional statistical evaluations to further assess the appropriateness of the resulting proposed SQVs. These additional statistical measures are believed to be less affected by the proportion of toxic and nontoxic samples in the data set. Further details of both reliability assessments can be found in the supplemental electronic files, as described in Section 1.4.

4.1 Standard Reliability Measures

The measures of reliability that were used to evaluate and select the final SQVs are defined and illustrated graphically in Figure 4-1:

- **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

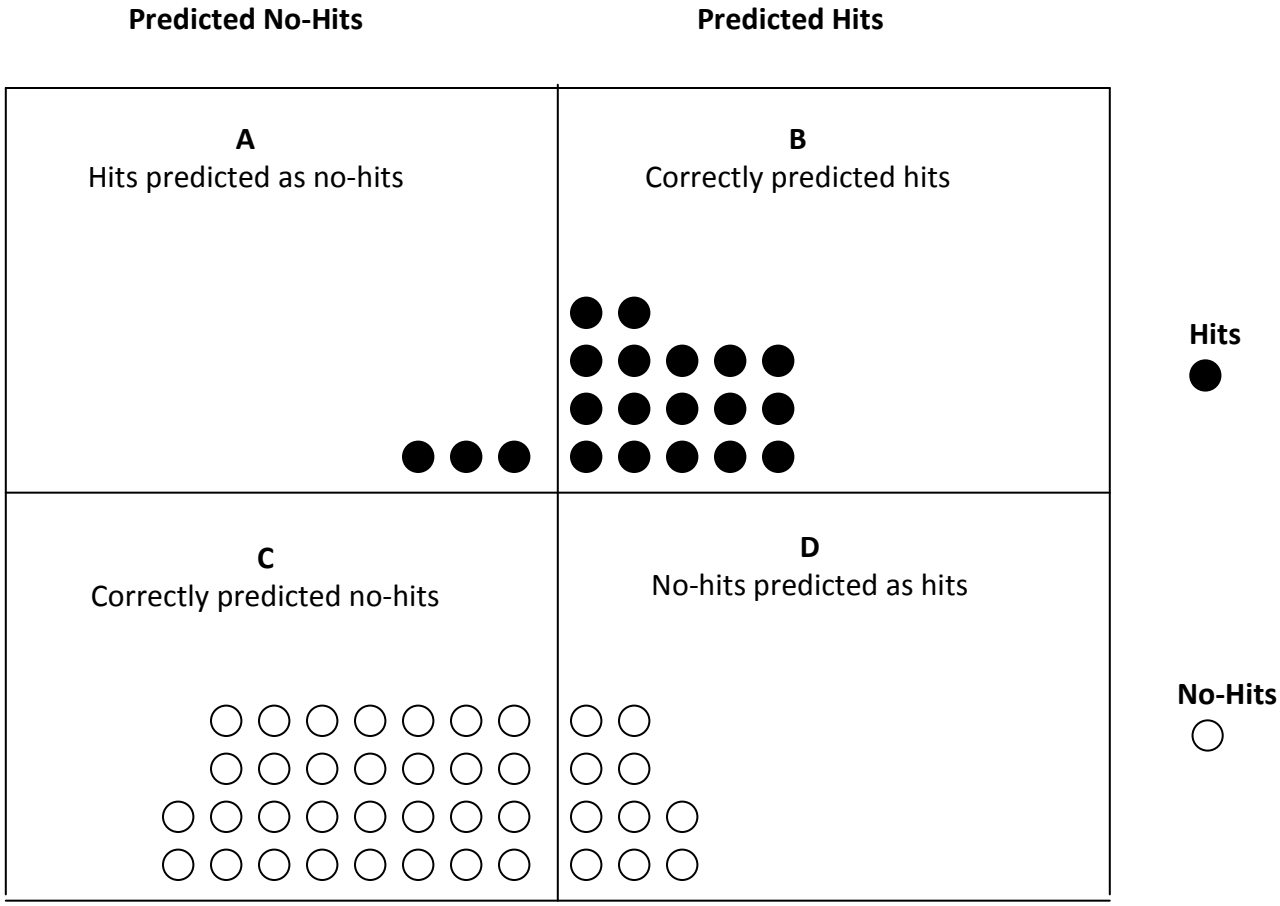
False positives and false negatives are the primary measures of predictive errors used in the reliability assessment. Each of the other reliability values is related to them in some way.

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 4-1. The goals were based on two factors: 1) the levels of error the agencies believed were appropriate for making regulatory decisions, and 2) the levels of reliability that were considered reasonably achievable based on previous results of the FPM model. The goals for the SQS/SL1 level were designed to be more protective by focusing on greater sensitivity (ability to correctly identify toxic sediments), while at the CSL/SL1 level, efficiency (ability to correctly identify clean sediments) to avoid unnecessary bioassay testing was considered equally important. Of the four measures, high predicted hit reliability (certainty that a predicted hit is actually a hit) is the hardest to achieve in a data set with mainly clean sediments, especially at the SQS/SL1 level. Therefore, that goal was also slightly lower than the others for the SQS/SL1 level.

Table 4-1. Reliability Goals for Proposed Freshwater SQVs

Reliability Measure	Goal (SQS/SL1)	Goal (CSL/SL2)
Sensitivity	80–90	75–85
Efficiency	70–80	75–85
Predicted hit reliability	70–80	75–85
Predicted no-hit reliability	80–90	75–85

Figure 4-1. Reliability Measures – Theoretical Example



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

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

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

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

Tables 4-2 and 4-3 show the reliability results for six different choices of false negative rates (0–30% at intervals of 5%) 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.

The cross-hatched box in each of the tables below indicates the row that was selected by the workgroup for derivation of the SQVs. The chemical concentrations corresponding with these rows appear in Tables 3-1 and 3-2. In each case, the selected rows met the reliability goals established by the workgroup. Therefore, the FPM values developed are considered appropriately sensitive, efficient, and reliable. Diagrams similar to Figure 4-1 showing correctly

and incorrectly predicted hits and no-hits are provided for each individual endpoint, as well as for the full set of proposed SQS/SL1 and CSL/SL2 values, in Figure 4-2 following the reliability tables. For the full SQG sets, only those stations that had at least three bioassay endpoints (two acute and one chronic or more) as described in the SMS were included, to avoid incorrectly identifying stations as nontoxic due to inclusion of historic data sets with less than a full suite of bioassays.

For consistency, and as a matter of policy, false negative rates for the individual bioassay endpoints were set at 20% for all endpoints except one. This row provides reasonable conservativism, given that all of these values are later combined and the lowest ones selected as the SQVs. In addition, the 20% row consistently met all of the workgroup's reliability goals, providing a good balance between false negatives and false positives and achieving high overall reliability for these bioassays. For one bioassay endpoint, *Hyalella* 10-day mortality at the SL2/CSL level, only the 25% false negative row met the workgroup's reliability goals for these three measures. In addition, it provided the best balance of false negatives and false positives, which is appropriate at the SL2/CSL level. Therefore, this row was selected for this bioassay endpoint.

It is important to note that a 20% false negative rate for a single endpoint at a station is not equivalent to an overall 20% false negative rate for that station. For each chemical, the SQVs for all of the bioassay endpoints were combined and the lowest values chosen as the SQS/SL1 and CSL/SL2 levels (see Section 5.2). Therefore, the regulatory levels for each chemical based on all of the available endpoints together will result in lower false negative rates than for any one bioassay endpoint alone. For further statistical evaluation of the level of bias and conservativism in the proposed SQVs, see Section 4.3.

In addition, multiple stations are used to make decisions about listing and cleaning up contaminated sites. With each additional station of data, the chances of missing a contaminated site decrease. For example, if the false negative rate is 20% for one station, or 0.2, then the false negative rate for three stations is 0.2 x 0.2 x 0.2, or 0.008, approximately 1%. In Ecology's Toxics Cleanup Program, three stations is currently the minimum number of data points required for making initial listing decisions, and many more stations are used for complete site evaluations.

While dredging decisions are often made on the basis of a single station representing a DMMU, bioassay testing is required for open-water disposal if sediment concentrations exceed the lower SQS/SL1 level, which is the lowest of all the endpoint concentrations for each chemical. The bioassay override procedures of the dredging program provide sufficient safety to ensure that unsuitable material is not disposed of in open water.

4.2 Comparison to Existing SQV Sets

Reliability tests were also run for other existing freshwater SQV sets to compare their predictive reliability for this updated 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).

For a detailed discussion of the narrative intent of these existing SQV sets, how each of them were calculated, the underlying data set used, the specific values used, and the original literature, please see SAIC and Avocet (2002). It should be noted that these SQV sets were calculated using different data sets from that used to calculated the FPM SQVs, as well as from each other. In addition, they include a variety of different bioassay endpoints, which are generally a subset of those used for the FPM, but may include some species that are regionally different from those used for the FPM. Finally, they are generally calculated on a combined-endpoint basis, while the FPM values are calculated (like the AET values) for individual endpoints. Nevertheless, these existing SQV sets are the only other alternatives available for regulatory use, so it is important to provide a comparison of reliability, subject to these caveats.

The reliability of the existing SQV sets for this data set was determined by entering the numerical values for each SQV set into the test row of the model calculation spreadsheets, and calculating the number of correct predictions of toxicity and non-toxicity, as well as false positives and false negatives.

The results are shown underneath each part of Tables 4-2 and 4-3 below for ease of comparison. The following observations can be made:

- 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 proposed 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 had at least twice the false positive rate of the FPM values, but often had twice the false negative rate 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 the upper and lower effects levels.

Table 4-2. 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	49.1	92.3	50.9	48.0	93.1	64.6
15	11.5	41.5	88.5	58.5	51.1	91.2	68.4
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	11.3	73.1	88.7	76.0	87.0	83.5

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 4-3. 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	4.5	40.1	95.5	59.9	24.2	99.0	64.1
10	9.0	36.9	91.0	63.1	24.8	98.1	66.4
15	14.9	25.7	85.1	74.3	30.6	97.4	75.5
20	20.9	20.0	79.1	80.0	34.6	96.6	79.9
25	23.9	18.0	76.1	82.0	36.2	96.3	81.3
30	29.9	12.4	70.1	87.6	43.1	95.6	85.6

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

Note: For this bioassay endpoint, the 25% false negative line was selected because it was the only line that met the reliability goals. In addition, this is a SL2/CSL endpoint; thus it is appropriate to maintain a balance between false negatives and false positives, with both being relatively low.

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

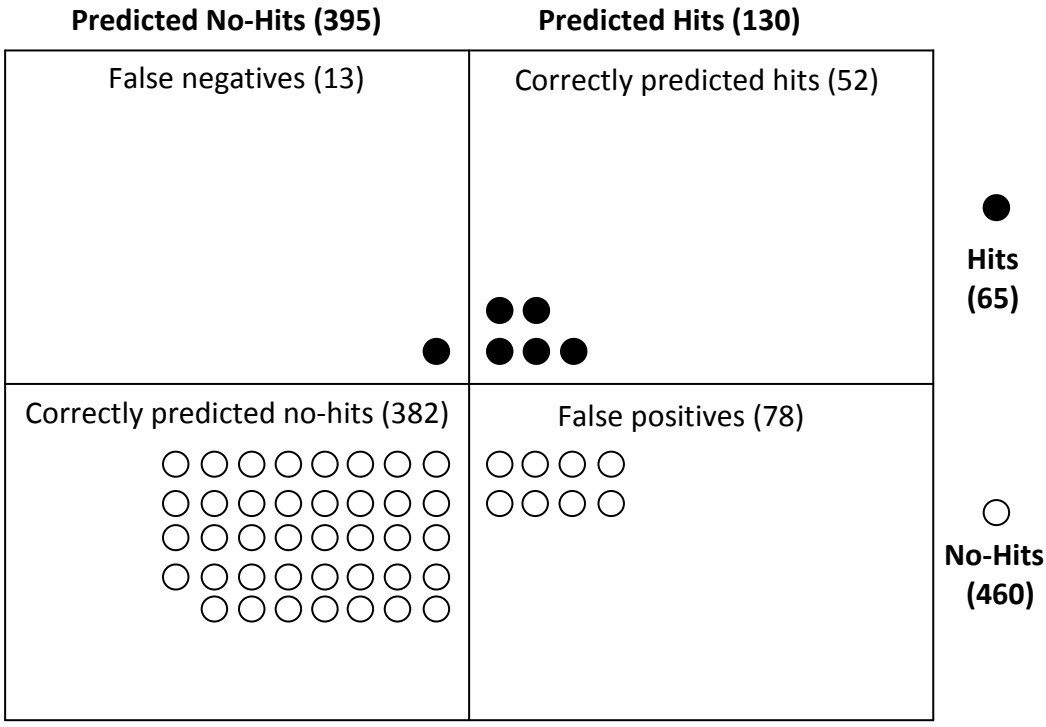
FPM FN Percentiles	% False Negatives	% False Positives	% Hit Reliability	% NoHit Reliability	% PredHit Reliability	%PredNoHit Reliability	% Overall Reliability
5	3.7	11.6	96.3	88.4	44.1	99.6	89.1
10	7.4	7.7	92.6	92.3	53.2	99.2	92.3
15	14.8	4.6	85.2	95.4	63.9	98.6	94.6
20	18.5	4.2	81.5	95.8	64.7	98.2	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

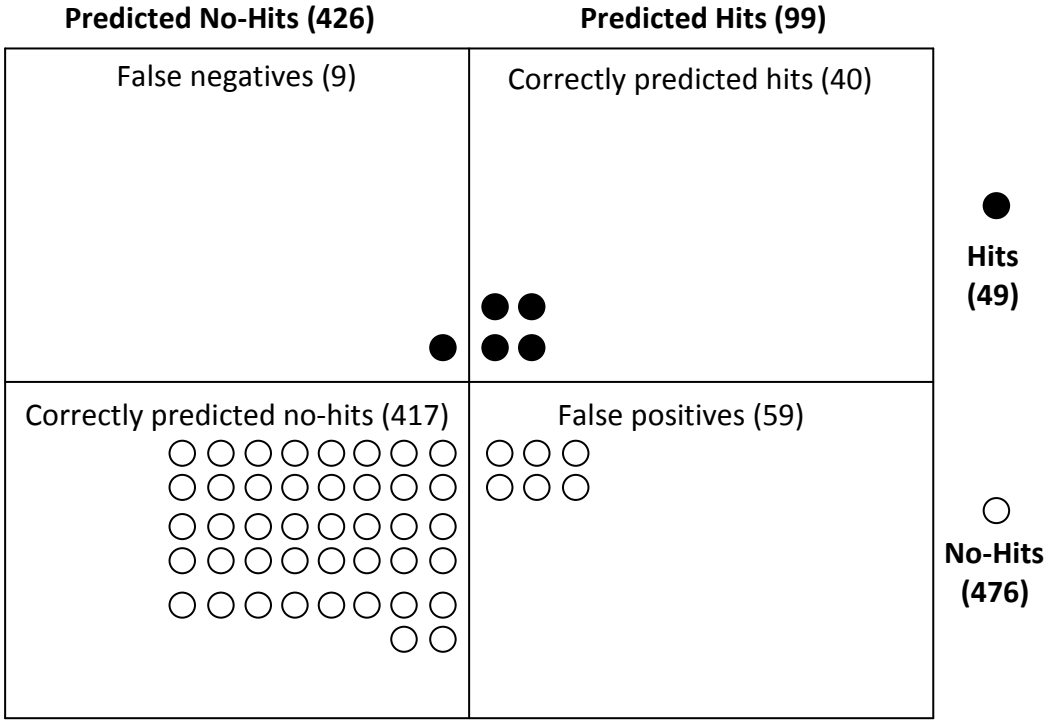
Figure 4-2. Predicted Hits and No-Hits vs. Actual Hits and No-Hits

Note: in all panels, each circle indicates approximately 10 stations

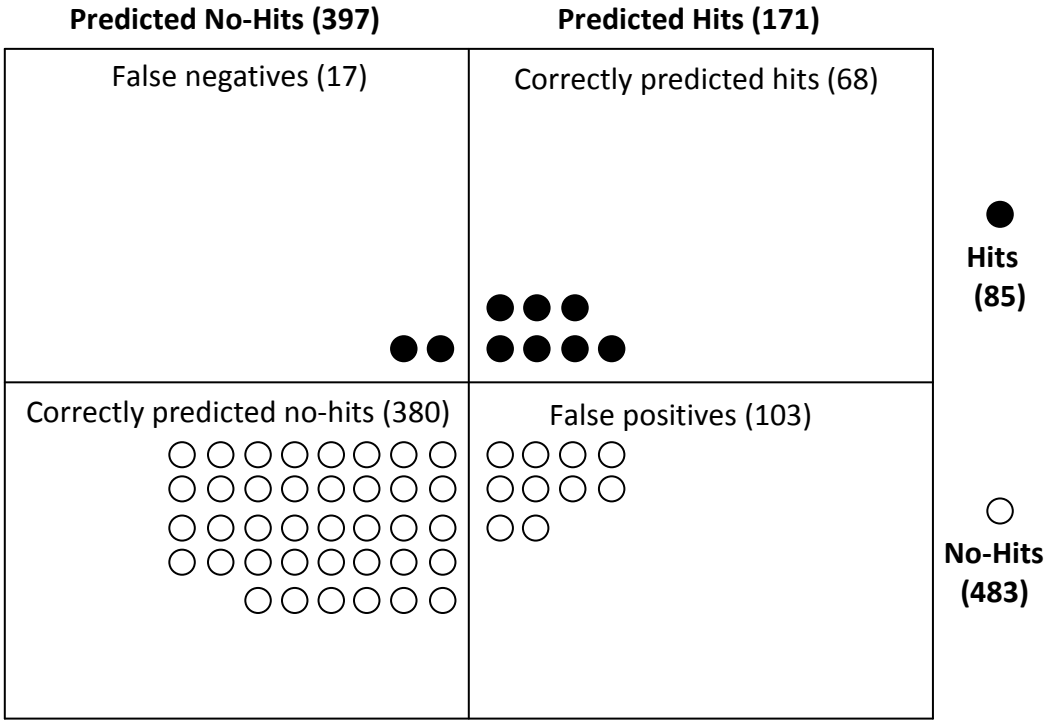
a) *Chironomus* 10-day growth SQS/SL1



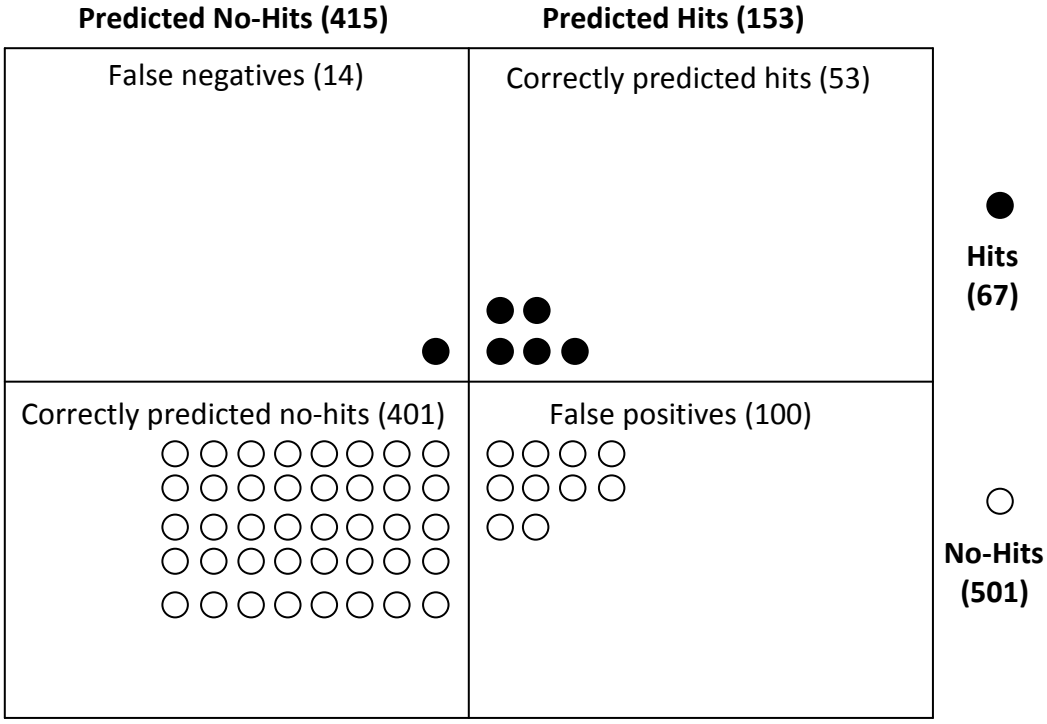
b) *Chironomus* 10-day growth CSL/SL2



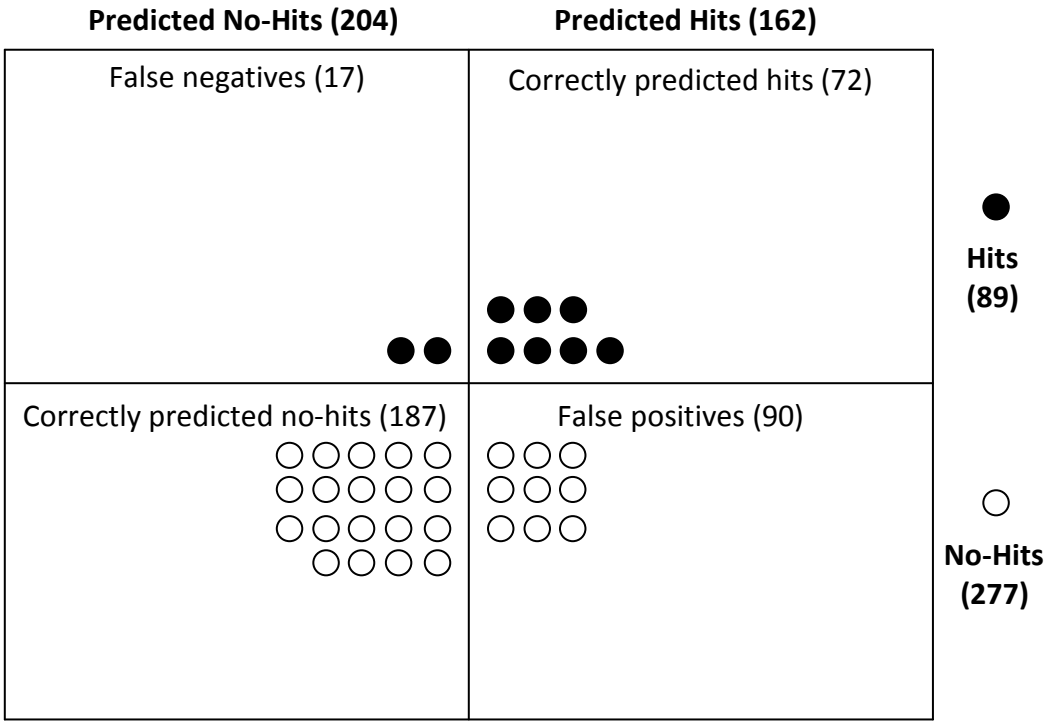
c) *Chironomus* 10-day mortality SQS/SL1



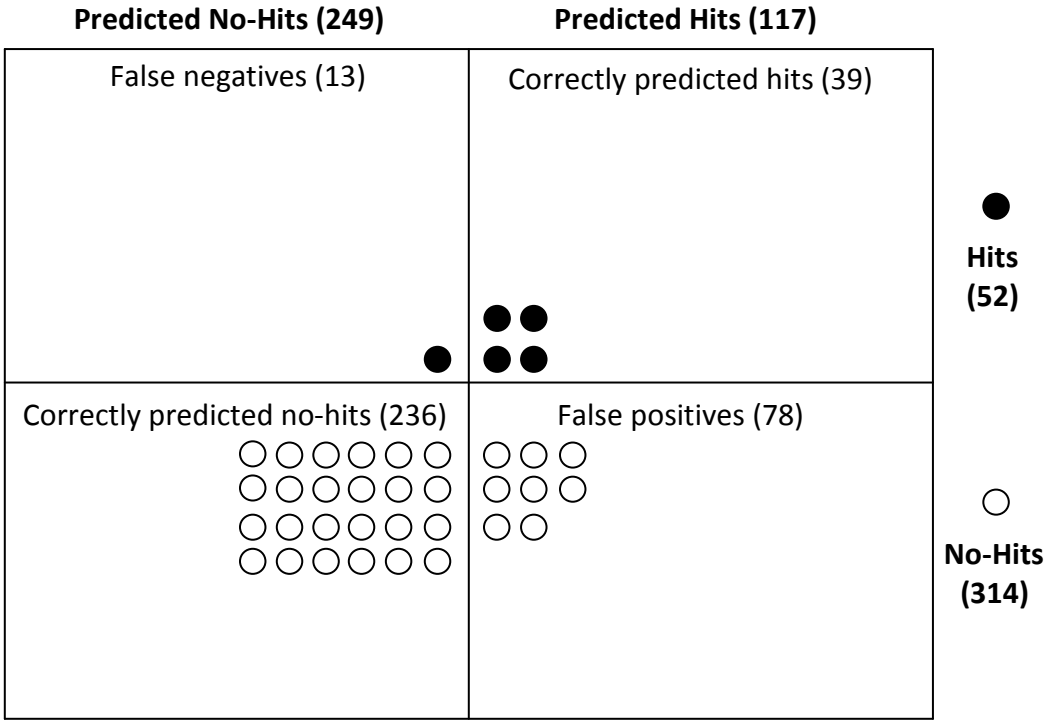
d) *Chironomus* 10-day mortality CSL/SL2



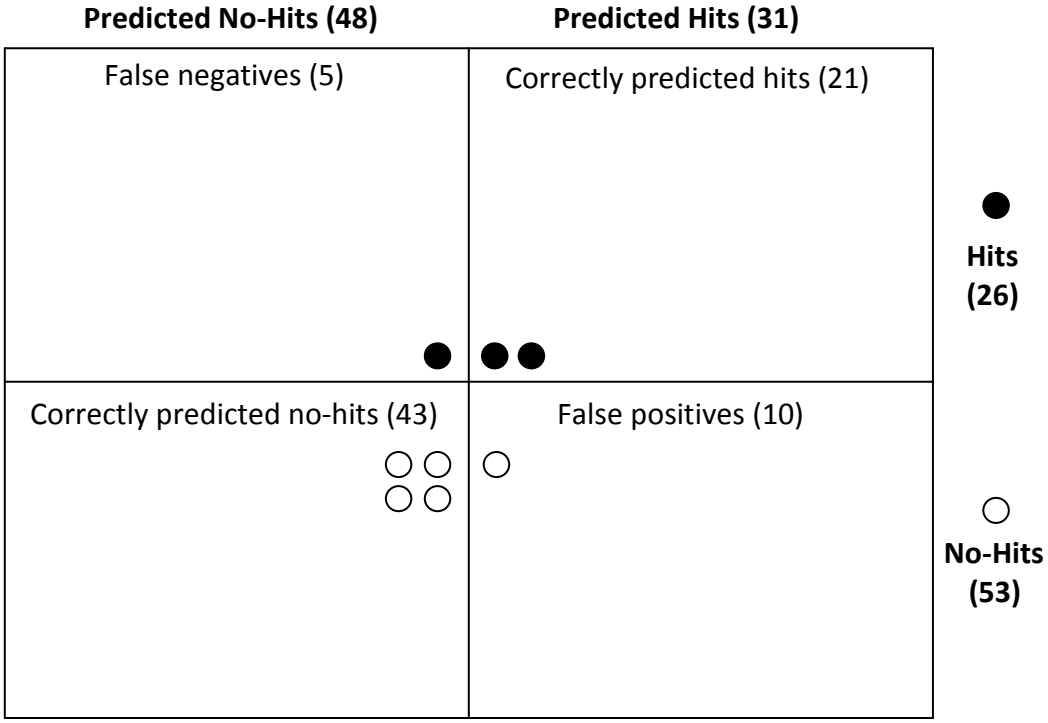
e) *Hyalella* 10-day mortality SQS/SL1



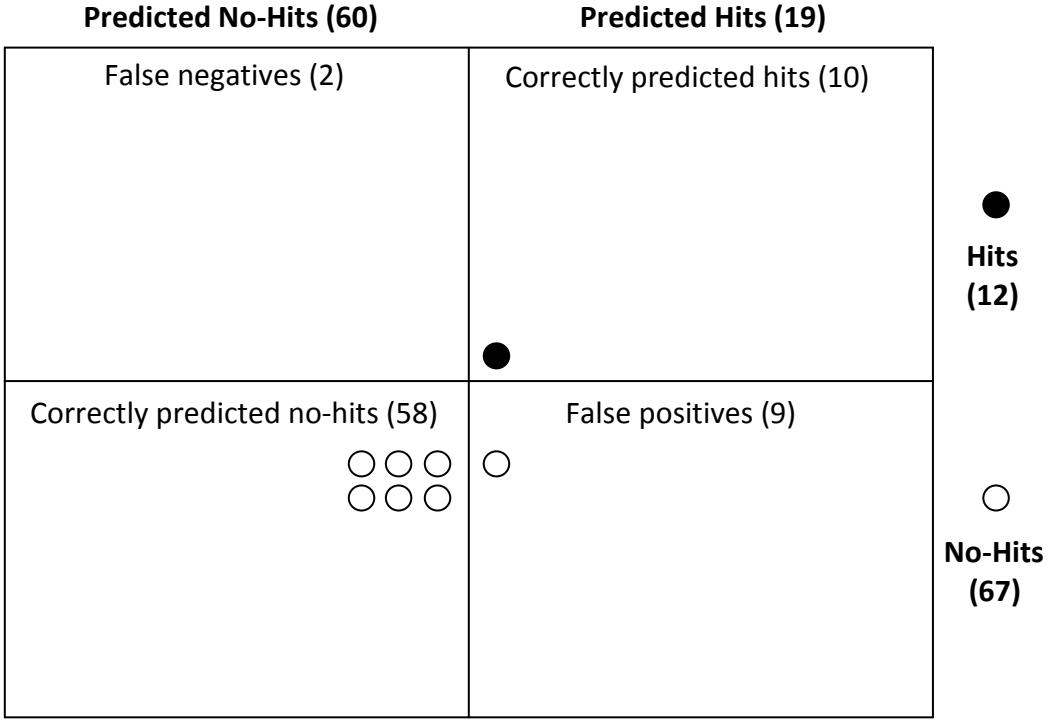
f) *Hyalella* 10-day mortality CSL/SL2



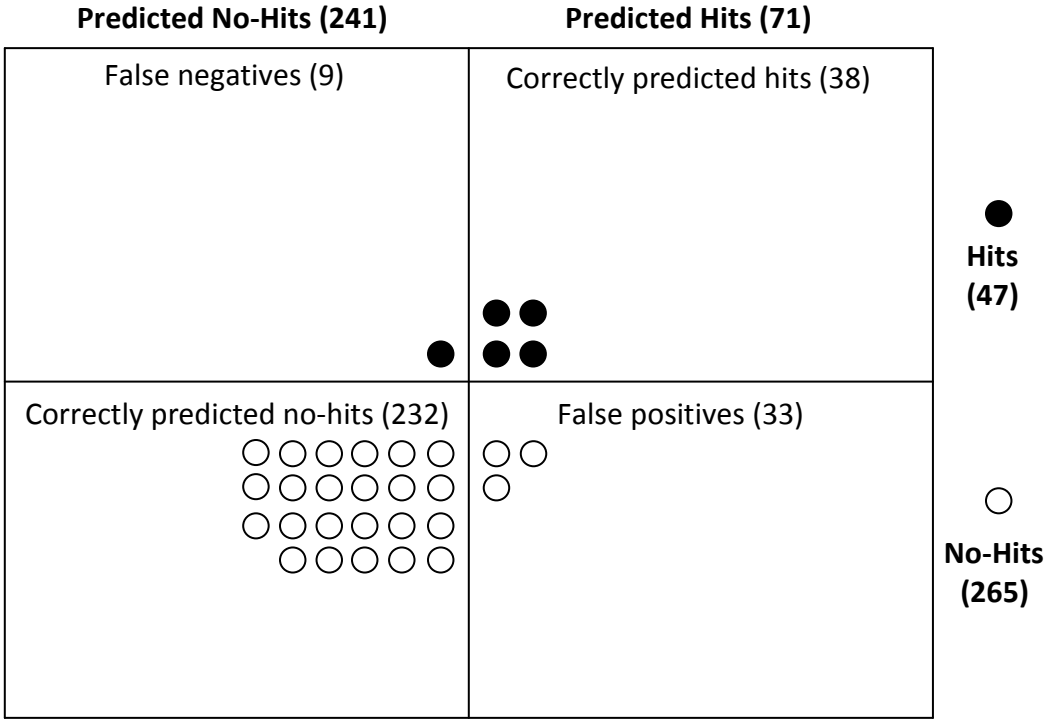
g) *Hyalella* 28-day growth SQS/SL1



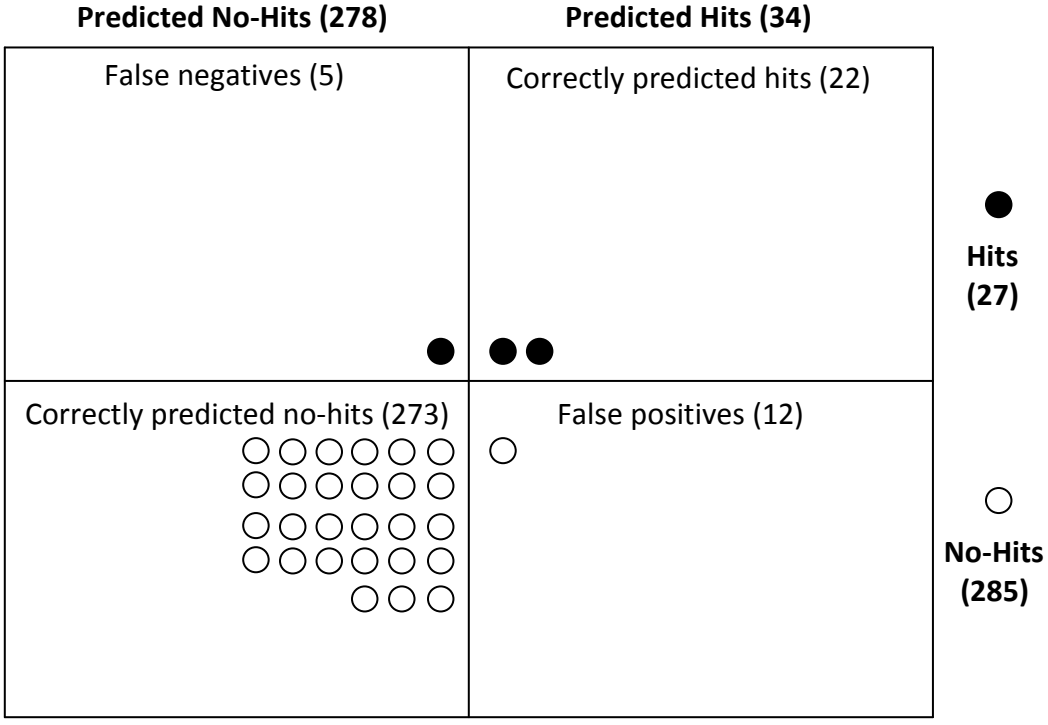
h) *Hyalella* 28-day growth CSL/SL2



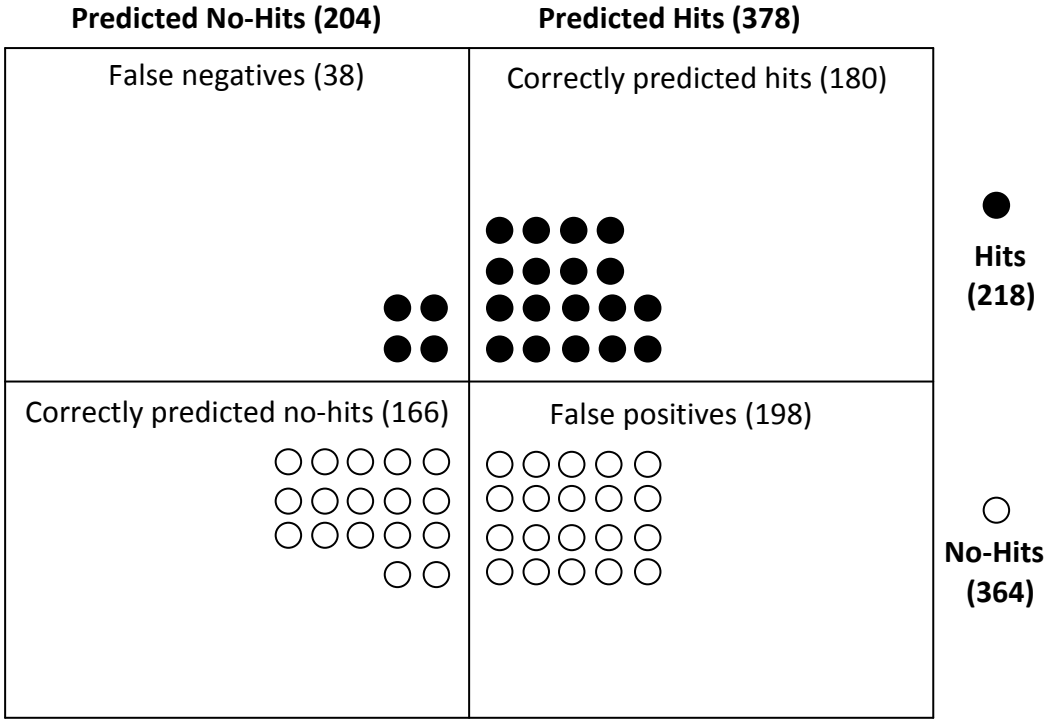
i) *Hyalella* 28-day mortality SQS/SL1



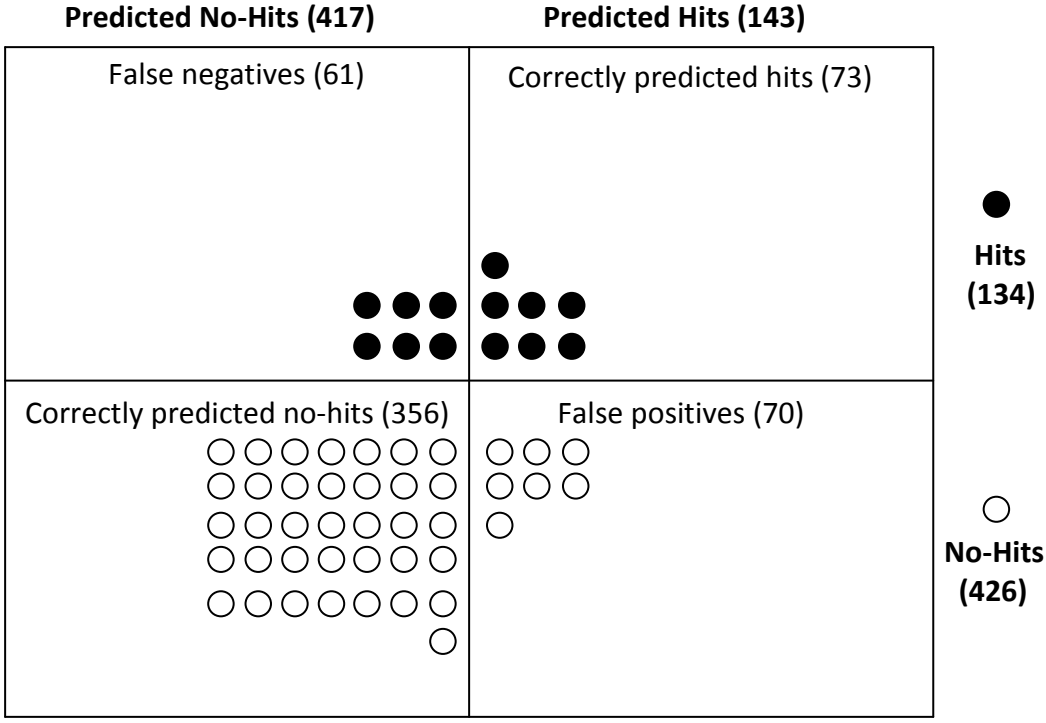
j) *Hyalella* 28-day mortality CSL/SL2



k) SQS/SL1 Proposed SQVs



l) CSL/SL2 Proposed SQVs



4.3 Supplemental Statistical Analyses

In addition to the standard reliability measures described above, EPA suggested that a variety of statistical measures be used that would be less affected or not affected by the prevalence of hits and no-hits in the data set. The following additional statistical measures were agreed upon between EPA and Ecology, all of which can also be calculated using the information in Figure 4-1, including:

- Bias
- Odds ratio
- Hanssen-Kuipers discriminant

Spreadsheets showing the calculation of these values are available as supplement electronic files, as described in Section 1.4. For each statistical measure, results are shown for both individual bioassay endpoints and the proposed SQS/SL1 and CSL/SL2 SQV sets. For statistical evaluation of the full SQG sets, only those stations that had at least three bioassay endpoints (two acute and one chronic or more) as described in the SMS were included, to avoid incorrectly identifying stations as nontoxic due to inclusion of historic data sets with less than a full suite of bioassays.

4.3.1 Bias

Bias is defined as the number of samples predicted to be toxic divided by the number of samples that are actually toxic. Thus, bias provides a simple measure of how protective a set of standards is:

- Bias > 1 indicates that the SQVs are protective and over-predict toxicity
- Bias = 1 indicates that the SQVs are appropriately predictive
- Bias < 1 indicates that the SQVs are under-protective and under-predict toxicity

Bias is calculated using the following formula, based on Figure 4-1: $(B + D)/(A + B)$.

Bias ranged from 1.2–2.3 for all individual endpoints, and from 1.1–1.7 for the draft SQVs (Tables 4-4 and 4-5). All chronic endpoints and the proposed SQVs had lower bias than all acute endpoints. The reason for this is not known, although it may be that the chronic tests have a larger number of true hits, which may lower the bias. The proposed SQVs had an even larger percentage of true hits than the overall data set due to exclusion of no-hit stations with only one or two bioassays, which also likely lowered the bias.

Ranges of bias were comparable for the SQS/SL1 and CSL/SL2 levels among individual endpoints. However, for the proposed SQVs, the bias was on the protective side (1.7) at the SQS/SL1 level and approximately 1 at the CSL/SL2 level. This suggests that the endpoints were combined appropriately in selecting the final criteria (see Section 3.3), erring on the protective side for the SQS/SL1 and achieving a good balance at the CSL/SL2 level.

Table 4-4. Bias at the SQS/SL1 Level

Endpoint	Correctly Predicted Hits	Correctly Predicted No-Hits	False Predicted Hits	False Predicted No-Hits	Bias
CH10G	52	382	78	13	2.0
CH10M	68	380	103	17	2.0
HY10M	72	187	90	17	1.8
HY28G	21	43	10	5	1.2
HY28M	38	232	33	9	1.5
Proposed SQVs	179	173	191	39	1.7

Table 4-5. Bias at the CSL/SL2 Level

Endpoint	Correctly Predicted Hits	Correctly Predicted No-Hits	False Predicted Hits	False Predicted No-Hits	Bias
CH10G	40	417	59	9	2.0
CH10M	54	375	126	13	2.3
HY10M	39	236	78	13	2.3
HY28G	10	58	9	2	1.6
HY28M	22	273	12	5	1.3
Proposed SQVs	47	352	74	60	1.1

4.3.2 Odds ratio

The odds ratio indicates the strength of a prediction, either for a given chemical or for a group of SQVs. The odds ratio is calculated as the likelihood that a prediction that a sample is toxic or nontoxic is correct over the likelihood that the prediction is incorrect. The odds ratio is calculated using the following equation based on Figure 4-1: (B + C)/(A + D). Thus, an odds ratio of 5 indicates that if there is a prediction of toxicity, the sample is 5 times more likely to actually be toxic than not. A higher odds ratio indicates stronger predictive capability.

The odds ratios range from 9–15 for acute mortality endpoints, and from 18–100 for chronic and growth endpoints (Table 4-6), suggesting that exceedance of SQVs for acute mortality endpoints is less likely to be predictive of true toxicity than exceedance of SQVs based on chronic or growth endpoints. Nevertheless, these odds ratios are relatively high for all individual endpoints, resulting in about a 1–10% chance of not seeing an effect when one is predicted.

The proposed SQVs have somewhat lower odds ratios, indicating that they are somewhat more protective than the values for the individual bioassays. The SQVs for both SQS/SL1 and CSL/SL2 levels have odds ratios of roughly 4:1–6:1, suggesting an 80–85% likelihood of exceeding the biological standards given a chemical SQV exceedance. These odds are in line with the policy goals used to calculate the guideline (80% overall accuracy, maximum of 20% false negatives and 20% false positives).

Table 4-6. Odds Ratios[a]

Endpoint	SQS/SL1	CSL/SL2
CH10G	20	31
CH10M	15	15
HY10M	8.8	9.1
HY28G	18	32
HY28M	30	100
Proposed SQVs	4.0	6.1

[a] The number of correct and false hits and no-hits are the same as shown in Tables 4-4 and 4-5.

4.3.3 Hanssen-Kuipers Discriminant

The Hanssen-Kuipers Discriminant is used to evaluate the fit of a model, and is frequently used to evaluate logistic regression models (but can be used for any model). The Hanssen-Kuipers Discriminant is a less general version of the Kappa statistic, and is used in cases where the prevalence of hits and no-hits in the data set is skewed. The Hanssen-Kuipers Discriminant is calculated using the following equation based on Figure 4-1: $((B \times C) - (A \times D))/((A + B) \times (C + D))$.

This statistic is believed to be unaffected by prevalence and ranges from 0–1. Like r^2, a Hanssen-Kuipers Discriminant value closer to 1 represents a better fit to the data. The following framework has been proposed in the epidemiological literature for interpreting model fit; however, this classification scheme is somewhat arbitrary and may or may not translate well to environmental data:

κ between .01–.20 = slight
κ between .21–.40 = fair
κ between .41–.60 = moderate
κ between .61–.80 = substantial
κ between .81–1 = nearly perfect

The results for the Hanssen-Kuipers Discriminant suggest models with "moderate" or "substantial" fits for each individual endpoint (Table 4-7). The best fits are again for the growth and chronic endpoints, with slightly lower values for the acute toxicity endpoints. The proposed SQVs have results suggestive of "fair" fits, with a somewhat better fit at the CSL/SL2 level. However, these represent combined SQVs that were not calculated using the FPM model, but were instead developed by the agencies through selecting the lowest or second-lowest of the individual endpoint values. Having not been developed using a modeling process, it may not be reasonable to expect the higher degrees of fitness to the data that the individual endpoints show. In addition, stations included in the proposed SQV assessments varied in the number and type of bioassays endpoints at each station, which may have reduced the fit.

Table 4-7. Hanssen-Kuipers Discriminants[a]

Endpoint	SQS/SL1	CSL/SL2
CH10G	0.63	0.69
CH10M	0.59	0.59
HY10M	0.48	0.50

HY28G	0.62	0.70
HY28M	0.68	0.77
Proposed SQVs	0.28	0.38

[a] The number of correct and false hits and no-hits are the same as shown in Tables 4-4 and 4-5.

This page was left blank intentionally

5. Selection of THE SQVs

5.1 Regulatory Considerations

Two effects levels were developed for each bioassay endpoint, one corresponding to the SQS/SL1 and one corresponding to the CSL/SL2. According to the statutory definition, SQS/SL1 represents a no acute or chronic adverse effects level and this is established as the minimum detectable difference from control, and CSL/SL2 represents a minor adverse effects level.

In the Washington State Sediment Management Standards, 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 the SQS to the CSL, 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.

For all dredging projects in the RSET program, the SL1 serves as the threshold above which biological testing is required to allow open water disposal, and below which open water disposal is permitted without biological confirmation.

As with the marine SQVs, the proposed freshwater SQVs were specifically developed 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 suite of bioassays and interpretive endpoints used to develop the SQVs will also be used to interpret the bioassay results to ensure consistency and maximize the reliability of the SQV predictions, although as additional freshwater bioassays are developed over time, the agencies may choose to apply them as appropriate.

The freshwater SQVs were developed to protect populations of benthic communities in sediments, given the wide natural variation in species abundance and richness seasonally and from year to year that exists, especially in freshwater systems. NOAA and USF&W were members of the RSET workgroup and accepted the task of determining whether the SQS/SL1 approach was protective of individual ESA-listed benthic species. NOAA and USF&W representatives reported to the workgroup that there were no listed benthic species in WA, OR, or ID that were present in areas where dredging or cleanup was likely to be conducted (personal communication to Keith Johnson, OR DEQ by Jeremy Buck, US F&W by e-mail, June 12,

2007). Therefore, lower values to protect individual ESA-listed benthic species were not developed.

5.2 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 associations between chemicals and toxicity for different endpoints. In addition, it reduces potential problems with combining historic toxicity data with variations among data sets in the bioassay endpoints and chemical analytes at each station, number and variability of replicates, etc.

However, 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, all of the values in Tables 3-1 and 3-2 were combined into a single distribution for each chemical from which the final SQVs would be selected. This distribution reflects the range of SQVs from the lowest no-effects level to the highest minor effects level. 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.

The following method was chosen by Ecology for setting the proposed SQVs:

- **SQS/SL1 –** Select the lowest value for each chemical.
- **CSL/SL2** – Select the next highest significantly different value (>20% higher than the SQS/SL1).

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. A 20% difference between the upper and lower values was chosen to reflect a typical analytical relative percent difference (RPD), and ensures that these values can be distinguished given the typical precision of available analytical methods. The degree of conservativism of these final values was evaluated in Section 4.3.1 and found to appropriately reflect Ecology's policy goals.

5.3 Proposed SQVs

The proposed SQVs based on the approach described above are shown in Table 5-1. For some chemicals, only an SQS/SL1 could be established; the remaining concentrations were all "greater than" values. 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 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.

The values in Table 5-1 are proposed SQVs, based on the many selections and method assumptions outlined in this report. Alternative choices could be made based on public and agency review that may 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, RSET, and the other agencies and programs that may choose to use these values, following appropriate public review and comment.

Table 5-1. Proposed Sediment Quality Values

Analyte	SQS/SL1	Source[a]	CSL/SL2	Source[a]
Conventional Pollutants (mg/kg)				
Ammonia	230	HY28M	300	HY28M
Total sulfides	39	CH10G	61	HY28M
Metals (mg/kg)				
Antimony[b]	0.3	HY10M	12	HY28M
Arsenic	14	HY28G	120	CH10G/CH10M
Cadmium	2.1	CH10M	5.4	HY28M
Chromium	72	HY28G	88	CH10G
Copper	400	HY28G	1200	HY28G
Lead	360	CH10G	> 1300	HY10M
Mercury	0.66	HY28G	0.8	CH10M/HY10M
Nickel	26	HY28G	110	CH10G
Selenium	11	HY28G	> 20	CH10G/HY28M
Silver	0.57	CH10G	1.7	HY28M
Zinc	3200	HY28G/HY28M	> 4200	HY10M
Organic Chemicals (µg/kg)				
4-Methylphenol	260	HY28M	2000	CH10M
Benzoic acid	2900	CH10M	3800	HY10M
beta-Hexachlorocyclohexane	7.2	CH10G	11	CH10M/HY28M
bis(2-Ethylhexyl)phthalate	500	HY10M	22000	HY10M
Carbazole	900	CH10M	1100	CH10M
Dibenzofuran	200	CH10G	680	CH10M/HY28M
Dibutyltin	910	CH10G/CH10M	130000	HY10M
Dieldrin	4.9	CH10G/CH10M	9.3	CH10M
Di-n-butyl phthalate	380	CH10G	1000	HY28M
Di-n-octyl phthalate	39	HY10M	> 1100	CH10G
Endrin ketone	8.5	CH10G/CH10M/HY28M	**	
Monobutyltin	540	CH10G/CH10M	> 4800	HY10M
Pentachlorophenol	1200	HY10M	> 1200	CH10G/CH10M
Phenol	120	HY28M	210	CH10M/HY28M
Tetrabutyltin	97	CH10G/CH10M	> 97	HY28M
Total DDDs	310	HY10M	860	CH10G
Total DDEs	21	HY10M	33	CH10M
Total DDTs	100	CH10M	8100	CH10M/HY28M
Total PAHs	17000	CH10G/HY10M	30000	CH10G
Total PCB Aroclors	110	HY10M	2500	HY10M
Tributyltin	47	HY10M	320	CH10M
Bulk Petroleum Hydrocarbons (mg/kg)				
TPH-Diesel	340	CH10M	510	CH10G/CH10M
TPH-Residual	3600	CH10M	4400	CH10G

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.
** No SQV could be set due to limited data above the SQS/SL1 concentration.
[a] 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.
[b] Not recommended for promulgation at this time; see Section 5.4.

5.4 Implementing the SQVs

The following information is provided to assist site managers and the regulated community in interpreting the values in Table 5-1, as well as to describe how to address chemicals not included in the table if found in sediments at a site or dredging project.

- **Chemicals not Included in Standard Analyte Lists.** For scientific or programmatic reasons, agencies may decide not to include all of the chemicals in Table 5-1 in their regulations, guidance, or standard analyte list. However, in that case, the values in this table provide useful guidance should one of these chemicals prove to be of concern for a specific site or project. At this time, Ecology is proposing not to include ***antimony*** in the SMS list, due to known issues with the analytical methods, a high level of false positives, and the SQS/SL1 value being below background. Removal of antimony affects correct identification of toxicity for only one station at the CSL level in the data set; thus, its removal is not expected to have a significant impact on identification of sites or dredged sediments with toxicity.

- **Background Concentrations.** The values in Table 5-1 can be considered risk-based values for the benthic community. However, the SMS and dredging guidance provide that if natural background concentrations are higher than the risk-based values, the background values may be used instead. Currently, Ecology is aware of one chemical on the list, ***antimony***, whose SQS/SL1 value may be below state-wide and/or local background concentrations. The Portland District Corps of Engineers has also reported that the SQS/SL1 value for ***nickel*** may be below background in some areas of Oregon. In specific areas such as those influenced by mining, regional geochemical concentrations for other metals may be higher than the SQS/SL1 values shown; this would need to be determined on a site-specific basis. Due to the modeling methodology used, none of the CSL/SL2 values are expected to be below background.

- **Practical Quantitation Limits (PQLs).** All detected concentrations above the method detection limit (MDL) were used for modeling. However, the SMS provides that the PQL will be used if it is higher than the risk-based value. At this time, Ecology is aware that the PQL may be higher than the SQS/SL1 for ***di-n-octyl phthalate*** and ***phenol***. As these SQS/SL1 values are higher than the MDL but lower than the PQL, the PQL may decline over time through analytical advances to below the risk-based value. Until then, the PQL should be used for regulatory decision-making at the SQS/SL1 level. None of the CSL/SL2 values are below the PQL.

- **Greater Than (>) Values.** As noted above, some chemicals have an SQS/SL1 value but only a "greater than" value at the CSL/SL2 level. These chemicals include ***lead, selenium, zinc, di-n-octyl phthalate, Endrin ketone, monobutyltin, pentachlorophenol,*** and ***tetrabutyltin***. Higher levels of effects may be observed above the "greater than" value than were present in this data set. Therefore, the "greater than" value has been included for site managers' information. At levels above those seen in this data set, bioassays should be run to identify the presence or absence of higher levels of adverse effects.

- **Chemicals of Low Concern for Benthic Toxicity.** The model also identified a number of analytes that were not associated with toxicity in the data set for any endpoint, or that had "greater than" values for all bioassay endpoints and effects levels. These chemicals are not considered of significant concern to benthic organisms within the concentration range found in the data set, and include: ***Aldrin, butyl benzyl phthalate, dimethyl phthalate, dioxins/furans, gamma-hexachlorocyclohexane, hexachlorobenzene, hexachloroethane, methoxychlor, retene, total chlordanes,*** and ***total Endosulfans***. The maximum concentrations of these chemicals observed in the data set are listed in Appendix B, Table B-3. Above the levels presented in Table B-3, the toxicity of these analytes is unknown, and bioassay tests should be run.

- **Nontoxic Analytes and Derived Quantities.** Several frequently reported analytes and derived quantities were not included in the modeling because they are not considered toxic chemicals under circumstances commonly encountered in sediments. However, there may be rare situations where these metals are toxic or contribute to toxicity (e.g., mining sites). These include ***grain size parameters, total solids, acid volatile sulfides, aluminum, beryllium, calcium, iron, magnesium, manganese, potassium, sodium,*** and ***vanadium***. ***TEQs*** of any kind are also derived quantities, usually associated with toxic mechanisms in vertebrates and calculated for higher trophic level risk assessments that do not apply to benthic organisms. These analytes and derived quantities generally do not present risks to the benthic community (exceptions are possible for highly concentrated waste materials).

- **Total Organic Carbon.** Although ***TOC*** is not itself an analyte of concern and was not included in the model, excessive TOC may cause high levels of ammonia and sulfides in sediments, which are analytes of concern, and/or may create an inappropriate substrate for benthic life if the TOC is anthropogenic in origin (Kendall and Michelsen 1997). In that case, the source of the high TOC would be treated as a deleterious substance or waste material.

- **Other Chemicals.** A variety of other chemicals have been analyzed in sediments but were not found at sufficient stations to warrant development of SQVs. If a chemical is not in Table 5-1 or on any of the lists above, it likely falls in this category. A complete list of chemicals with <30 detections is listed in Appendix B, Table B-1. If a chemical is not found on any of the lists above or in Table B-1, it was either never analyzed for or never detected in the data set. If a site or dredging project includes frequent detections or high levels of any such chemicals, bioassay tests should be run to evaluate their toxicity.

- **Applicability to Unique Sites.** There are sites where unique geochemical conditions warrant initial testing using bioassays. While the SQVs are developed from data representative of the majority of freshwater sediment sites encountered in the northwest, it is recognized that benthic toxicity at sites with unique geochemical characteristics will differ and the SQVs are not representative of those sites (e.g., bogs, alpine wetlands, sites with mining, milling or smelting activities, substantial waste deposits, or sites with

unique pH, alkalinity, or other geochemical characteristics). Freshwater bioassays should be used to assess toxicity under these conditions.

6. 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 three acute and two 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 results in endpoint-specific SQVs with a sensitivity of 75-80%, efficiency of 65-95%, and overall reliability of 70-85%, depending on the specific endpoint and effects level. Additional statistical analyses confirmed that the SQS/SL1s were appropriately, but not unreasonably, biased on the protective side, and that the CSL/SL2s were evenly balanced between false positives and false negatives. The models for the individual endpoints were found to have a good fit to the data.

- **Comparison to Existing SQVs.** Compared to other SQV sets available for use, 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 more sensitive than the existing SQV sets.

- **Recommended SQVs.** Based on the conclusions above and the results of the reliability and statistical analyses, SQVs for both the SQS/SL1 and the CSL/SL2 levels are proposed for public review and adoption. 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. The method also allows site-specific values to be calculated for unusual or large sites.

- **Benthic Toxicity Only.** 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.

- **Additional Information for Site Managers.** Additional information on how to implement these values and considerations for sites with unique geochemistry is included in Section 5.4 and Appendix B, including lists of chemicals that were screened out and the reasons for doing so, and how to evaluate chemicals that do not have recommended SQVs.

This page was left blank intentionally

7. 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, Mach C, Parrett K. 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, Sinnege TL, Seinen W, Hermens JLM. 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.

DMMP. 2009. Dredged Material Evaluation and Disposal Procedures (User's Manual). Prepared by the Dredged Material Management Office, U.S. Army Corps of Engineers, Seattle WA on behalf of the Dredged Material Management Program agencies.

Ecology. 1989. Data Validation Guidance Manual for Selected Sediment Variables. June 1989. Prepared by PTI Environmental Services, Bellevue, WA for Washington Department of Ecology, Sediment Management Unit, 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.

Field LJ, Norton SB, MacDonald DD, Severn CG, Ingersoll CG. 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 DF, Gustafson DA, Shaw TC. 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 RO. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York, NY.

Hermens J, Canton H, Janssen P, de Jong R. 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, Brockhuyzen E, Canton H, Wegman R. 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, Leeuwangh P, Musch A. 1985b. Joint toxicity of mixtures of groups of organic aquatic pollutants to the guppy (*Poecilia reticulata*). *Ecotox. Environ. Safety* 9:321–326.

Ingersoll CG, Ivey CD, Kemble NE, Mount DR, Field J, MacDonald DD, Smorong D. 2008. 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*). 2008 SETAC National Meeting, Tampa, FL.

Kendall D, Michelsen TC. 1997. Management of Wood Waste under Dredged Material Management Programs (DMMP) and the Sediment Management Standards (SMS) Cleanup Program. SMS Technical Information Memorandum, Washington Department of Ecology, Olympia, WA.

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

Michelsen TC, Shaw TC. 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.

SEF. 2009. Sediment Evaluation Framework for the Pacific Northwest. Interim Final. September 30, 2006. Prepared by the US Army Corps of Engineers (Seattle District, Portland District, Walla

Walla District, and Northwestern Division); U.S. Environmental Protection Agency Region 10; Washington Department of Ecology; Washington Department of Natural Resources; Oregon Department of Environmental Quality; Idaho Department of Environmental Quality; National Marine Fisheries Service; and U.S. Fish and Wildlife Service.

Sokal RR, Rohlf FJ. 1981. Biometry. Second Edition. W.H. Freeman and Company, San Francisco, CA.

Stirling SK, RSET (Regional Sediment Evaluation Team). 2008. Reference Areas for Freshwater Bioassays. Dredged Material Management Program (DMMP) Clarification Paper, Final.

Swartz RC, Schults DW, Ozretich RJ, Lamberson JO, Cole FA, DeWitt TH, Redmond MS, Ferraro SP. 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.

US EPA. 1986. Recommended Protocols for Measuring Conventional Sediment Variables in Puget Sound. March 1986. Prepared by the Puget Sound Water Quality Action Team, Olympia, WA for the U.S. Environmental Protection Agency, Region 10, Seattle, WA.

US EPA. 1997a. Recommended Protocols for Measuring Metals in Puget Sound Marine Water, Sediment, and Tissue Samples. Prepared by Puget Sound Water Quality Action Team, Olympia, WA for the U.S. Environmental Protection Agency, Region 10, Seattle, WA.

US EPA. 1997b. Recommended Protocols for Measuring Organic Compounds in Puget Sound Water, Sediment, and Tissue Samples. Prepared by the Puget Sound Water Quality Action Team, Olympia, WA for the U.S. Environmental Protection Agency, Region 10, Seattle, WA.

US EPA. 1997c. Recommended quality assurance and quality control guidelines for the collection of environmental data in Puget Sound. Prepared by the Puget Sound Water Quality Action Team, Olympia, WA for the U.S. Environmental Protection Agency, Region 10, Seattle, WA.

US EPA. 1999. USEPA Contract Laboratory Program National Functional Guidelines for Organic Data Review. EPA/540/R-99/008. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, Washington, DC.

US 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, Washington, DC.

US EPA. 2004. USEPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review. Final. OSWER 9240.1-45. EPA 540-R-04-004. U.S. Environmental Protection Agency, Office of Superfund Remediation and Technology Innovation (OSRTI), Washington, DC.

US EPA 2007. SW-846 on-line. Test methods for evaluating solid wastes, physical/chemical methods. www.epa.gov/epaoswer/hazwaste/test/main.htm. Accessed on July 17, 2006. U.S. Environmental Protection Agency, Office of Solid Waste, Washington, DC.

APPENDIX A. LIST OF SURVEYS

		Bioassay Endpoints[b]					Analyte Classes[c]								
State (E/W)[a]	Survey	CH10G	CH10M	HY10M	HY28G	HY28M	CON	MET	SV	VOL	CL	PP	DF	TPH	Reference
OR (W)	CBSLOUGH	0	0	20	0	0	X	X	X			X			Columbia Slough Sediment Analyses and Remediation Project, Phase 1 Report, Dames & Moore for City of Portland, 1991
OR (W)	FWDMMP05	26	26	26	0	0	X	X	X		X	X			Sediment Characterization Report, Lower Willamette River Federal Navigational Channel, Corps of Engineers, 2005
OR (W)	FWJSLK04	8	8	8	0	0	X	X	X			X		X	Johnson Lake Site Investigation Report, Arcadis for Owens-Brockway Glass Container, Inc., 2004
OR (W)	FWPHBR04	227	233	0	0	233	X	X	X	X	X	X	X	X	Portland Harbor Remedial Investigation Round 2 Data, Lower Willamette Group, 2004
OR (W)	FWTEKX07	13	13	13	0	0	X	X	X						Tektronix Site Remedial Investigation, Phase III, Windward Environmental, 2007
OR (W)	FWWRSD04	21	21	21	21	21	X	X	X			X			Willamette River Federal Navigation Channel O&M Sediment Characterization Report, Corps of Engineers, 2004
OR/WA (W)	LCBWRS93	0	0	15	0	0	X	X	X			X			Lower Columbia River Backwater Reconnaissance Survey, TetraTech for Lower Columbia River Bi-State Program, 1994
OR (W)	MBCREOS3	43	43	43	0	0	X	X	X						McCormick & Baxter RD Phase I Sediment Survey, Oregon DEQ, 2002
OR (W)	MBCREOS4	17	18	18	0	0	X	X	X						McCormick & Baxter RD Phase II Sediment Survey, Oregon DEQ, 2002
OR (W)	PPTLDT24	4	4	4	0	0	X	X	X			X			Sediment Characterization Study, Marine Terminal 2 Berths 203-206 and Marine Terminal 4 Berth 416, Hart Crowser for Port of Portland, 1999
OR (W)	PSYD&M97	0	0	3	0	0	X	X	X			X			Portland Shipyard Environmental Audit, Dames & Moore for Cascade General, 1998
OR (W)	PSYSEA98	55	55	55	0	0	X	X	X			X			Portland Shipyard Sediment Investigation Data Report, Striplin Env. Assts. for Port of Portland, 1998
OR (W)	ROSSIS99	11	11	11	0	0	X	X	X	X	X	X		X	Ross Island Facility Site Investigation, Hart Crowser for Port of Portland, 2000
OR (W)	TOSCO99	2	2	2	0	0	X	X	X			X			TOSCO Portland Terminal, 1999 Sediment Sampling Results, Portland District Corps of Engineers, 1999
OR (W)	WILREF02	3	3	3	0	0	X	X	X	X		X		X	Willamette Reference Survey, Hart Crowser for the Portland District Corps of Engineers, 2002
OR (W)	WLRPT498	18	18	18	0	0	X	X	X			X			Terminal 4 Slip 3 Sediment Investigation, Hart Crowser for Port of Portland, 1998
OR (W)	WRD&M98	0	0	2	0	0	X	X	X						Portland Shipyard Environmental Audit, Dames & Moore for Cascade General, 1998
WA (E)	BOISECAS	0	0	4	0	0	X	X	X						Class II Inspection of the Boise Cascade Pulp and Paper Mill Wallula Washington, WA Dept. of Ecology EILS, 1993
WA (E)	FWSPOR00	0	0	0	8	8	X	X	X			X			Chemical Analysis and Toxicity Testing of Spokane River Sediments Collected in October 2000, WA Dept. of Ecology EAP, 2001
WA (E)	FWUPCR05	50	50	0	50	50	X	X	X			X	X		Upper Columbia River Site CERCLA RI/FS, CH2M Hill for US EPA Region 10, 2005
WA (E)	SPOKNR94	0	0	3	0	0	X	X	X						Spokane River PCB Study, WA Dept of Ecology EILS, 1994
WA (W)	CARGIL01	0	3	3	0	0	X	X	X			X			Cargill Irving Elevator Terminal, Cargill Irving, 2001
WA (W)	CEDARIV	0	0	5	0	0	X	X	X						Sediment Sampling and Analysis Report Cedar River Delta Sediments, Golder Assts. for City of Renton, 1992
WA (W)	FWLKUN01	5	4	4	0	0	X	X	X		X	X		X	Lake Union Sediment Study, King County DNR, 2001
WA (W)	LKUNDRDK	0	0	4	0	0	X	X	X			X			Sediment Monitoring Program Results Lake Union Drydock Company, Hart Crowser, 1992
WA (W)	LKUNION	0	0	9	0	0	X	X	X			X			Survey of Contaminants in Lake Union and Adjoining Waters, WA Dept. of Ecology EILS, 1989
WA (W)	LKWA00	0	28	28	0	0	X	X	X			X			Lake Washington Baseline Sediment Study, King County, 2000
WA (W)	LUUCSO00	0	6	6	0	0	X	X	X			X			Lake Union University Regulator CSO Post Separation Study, King County, 2000
WA (W)	QUEBAX1	0	0	4	0	0	X	X	X						Distribution and Significance of PAHs in Lake Washington Sediments Adjacent to Quendall Terminals, WA Dept. of Ecology EILS, 1991
WA (W)	QUEBAX3	0	0	3	0	0	X	X	X						Results of Sediment Sampling in the JH Baxter Cove Lake Washington, WA Dept. of Ecology EILS, 1992
WA (W)	SALIII97	22	22	22	0	0	X	X	X			X			Salmon Bay Results of Phase III Sampling, WA Dept of Ecology EAP, 2000
WA (W)	SEACOM94	0	0	3	0	0	X	X	X			X			Sediment Sampling Report Seattle Commons Parcel C Seattle, Washington, 1994
WA (W)	TRI-STAR	0	0	3	0	0	X	X	X			X			Tri-Star Marine NPDES Sediment Monitoring, Beak Consultants, 1997
WA (W)	WEYLONG	0	0	3	0	0	X	X	X						Class II Inspection of Weyerhaeuser Longview Pulp and Paper Mill, WA Dept. of Ecology EILS, 1991
		525	**568**	**366**	**79**	**312**									

[a] OR = Oregon, WA = Washington, E = east of the Cascade Mountains, W = west of the Cascade Mountains.
[b] 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.
[c] CON = conventionals, MET = metals, SV = semivolatiles, CL = chlorinated hydrocarbons, TPH = total petroleum hydrocarbons, VOL = volatiles, PP = pesticides/herbicides, polychlorinated biphenyls (Aroclors), DF = dioxins/furans

APPENDIX B

DATA SCREENING

Section 2.2 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, stations, and chemical and biological data that were screened out of the data set.

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 different sensitivity to contaminants. Although for some time there had been a general sense that the early McCormick & Baxter results were unusual, this was 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 substantially 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 (Burt Shepard, US EPA Region 10, personal communication to T. Michelsen, Avocet Consulting on 15 April 2011). 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 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. 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
- **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 that 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)

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 such a large 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	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
Acetone	30*
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
Methylene chloride	1
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 had >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 PCB Aroclors:** 1016, 1221, 1242, 1248, 1254, 1260, 1268 (no congener data were available)

ANOVA Screening

The second step of the model runs is to evaluate which chemicals are associated with toxicity in the data set for each chemical and each endpoint (Table B-2). This evaluation is described in Section 2.6, and electronic spreadsheets showing the basis and detailed results of this screening are available as described in Section 1.4.

As a result of this evaluation, it was determined that the following chemicals had no association with 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 are not associated with toxicity to the benthic community at sediment concentrations historically observed in the environment, and thus, SQVs do not need to be set for them.

In addition to these chemicals, some chemicals were not associated with 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 associated with 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	CH10M SQS/SL1	CH10M CSL/SL2	CH10G SQS/SL1	CH10G CSL/SL2	HY10M SQS/SL1	HY10M CSL/SL2	HY28M SQS/SL1	HYA28M CSL/SL2	HY28G SQS/SL1	HY28G 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
Mercury	1*	1**	1	1*	0	0	0*	1	0	0

Analyte	CH10M SQS/SL1	CH10M CSL/SL2	CH10G SQS/SL1	CH10G CSL/SL2	HY10M SQS/SL1	HY10M CSL/SL2	HY28M SQS/SL1	HYA28M CSL/SL2	HY28G SQS/SL1	HY28G CSL/SL2
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 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

Chemical Analyte	Reason for Screening	Maximum concentration for benthic organisms[a]
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

Chemical Analyte	Reason for Screening	Maximum concentration for benthic organisms[a]
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

Chemical Analyte	Reason for Screening	Maximum concentration for benthic organisms[a]
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

Chemical Analyte	Reason for Screening	Maximum concentration for benthic organisms[a]
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] Concentration below which no association with toxicity was observed in the data set used to calculate the SQVs. Does not address potential bioaccumulation toxicity to wildlife, fish, or humans.

This page was left blank intentionally

APPENDIX C

SELECTION OF BIOLOGICAL EFFECTS LEVELS

This appendix provides a detailed discussion of the selection of biological effects levels for each bioassay endpoint at both the SQS/SL1 and CSL/SL2 effects levels. Table 2-2 in the main report presents the parameters described below.

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 (ASTM 2005) 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 suggested 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 WA SMS 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%. Within this range, statistical testing of commercial data from WA and OR determined that correlations between hit stations and toxicity improved at a threshold of 15% and did not increase substantially thereafter; thus, the 15% level was selected. Therefore, a detectable difference could occur 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 thresholds should be statistically significant nearly all of the time, with the minimum detectable difference occasionally exceeding the SQS/SL1 numeric threshold, but not likely exceeding the CSL/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.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 2–20%, with a mean of 8%. Therefore, a detectable difference could occur 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 statistically significant most of the time, with the minimum detectable difference at times exceeding the SQS/SL1 numeric threshold, but not likely exceeding the CSL/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 16–50%. 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 statistically significant about half the time, and the numeric level corresponding to the CSL/SL2 should be statistically significant about 80% of the time.

***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 suggested 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 2–12%, with a mean of 8%. However, statistical testing of commercial data from WA and OR determined that correlations between hit stations and toxicity improved at a threshold of 20% and did not increase substantially thereafter; thus, the 20% level was selected. Therefore, a detectable difference could occur at levels as low as 20% mortality, ranging in the worst case up to about 40% 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 statistically significant most 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 5–24%. 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 statistically significant well over half of the time, and the CSL/SL2 levels should be statistically significant 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.

This page was left blank intentionally

APPENDIX D

INTERIM MODEL RUNS

This appendix provides the results and discussion of interim model runs used to develop the final data set and modeling approach. In all cases, only the reliability results were calculated and presented to the workgroup, to preclude bias associated with the numeric results for individual chemicals. The reliability results are also presented here and an explanation is provided of the question being addressed and what decisions were made based on the results.

All of the results presented in this appendix are based on early versions of the data set, which did not have the same level of reliability as the final data set. As each quality assurance or database issue was worked through, the data set and/or the modeling approach incrementally improved, until the reliability goals were ultimately reached for each bioassay endpoint and effects level. Therefore, the results presented here are only for comparative purposes to illustrate the decisions that were made at the time.

D.1. Petroleum Hydrocarbons

The model was run using 1) total PAHs, 2) TPH-diesel and TPH-residual, and 3) both combined for two different data sets:

- The large data set included all data in the database, for which all stations had PAH data but only about 1/3 had TPH data. Because TPH alone could not predict toxicity in the other 2/3 of the stations, only PAH alone was compared with PAH + TPH combined.
- The small data set included only those stations that had both PAH and TPH data. For this data set, PAH alone, TPH alone, and PAH + TPH combined were compared.

For this modeling exercise, two representative bioassay endpoints were selected, Chironomus 10-day growth and Hyalella 28-day mortality. The results are shown in Table D-1 below.

For the large data set, the reliability was always improved by adding TPH over PAH alone, even though there were TPH data for only 1/3 of the stations. For the small data set, TPH was much more reliable than PAH alone, with very similar performance between TPH + PAH and TPH alone. Based on these results, it was agreed that both TPH and total PAHs would be used for the modeling, and that TPH should be analyzed more frequently at sediment sites where bulk petroleum may be an issue.

Table D-1. TPH vs. PAH Comparisons

a. CH10G Small Data Set SQS/SL1

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	42.6	100.0	57.4	24.8	100.0	62.7
5	2.9	36.1	97.1	63.9	27.4	99.4	68.0
10	8.6	28.5	91.4	71.5	31.1	98.3	73.9
15	14.3	21.3	85.7	78.7	36.1	97.5	79.6
20	20.0	16.5	80.0	83.5	40.6	96.7	83.1
25	22.9	13.3	77.1	86.7	45.0	96.4	85.6
30	28.6	5.6	71.4	94.4	64.1	95.9	91.5

TPH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	34.1	100.0	65.9	29.2	100.0	70.1
5	2.9	26.5	97.1	73.5	34.0	99.5	76.4
10	8.6	17.3	91.4	82.7	42.7	98.6	83.8
15	14.3	12.4	85.7	87.6	49.2	97.8	87.3
20	20.0	12.4	80.0	87.6	47.5	96.9	86.6
25	22.9	9.2	77.1	90.8	54.0	96.6	89.1
30	28.6	4.8	71.4	95.2	67.6	96.0	92.3

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	31.3	100.0	68.7	31.0	100.0	72.5
5	2.9	23.7	97.1	76.3	36.6	99.5	78.9
10	8.6	14.9	91.4	85.1	46.4	98.6	85.9
15	14.3	10.4	85.7	89.6	53.6	97.8	89.1
20	20.0	10.4	80.0	89.6	51.9	97.0	88.4
25	22.9	10.4	77.1	89.6	50.9	96.5	88.0
30	28.6	6.0	71.4	94.0	62.5	95.9	91.2

b. CH10G Small Data Set CSL/SL2

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	44.2	100.0	55.8	18.6	100.0	59.9
5	3.8	34.9	96.2	65.1	21.7	99.4	68.0
10	7.7	31.4	92.3	68.6	22.9	98.9	70.8
15	11.5	21.3	88.5	78.7	29.5	98.5	79.6
20	19.2	11.2	80.8	88.8	42.0	97.9	88.0
25	23.1	7.0	76.9	93.0	52.6	97.6	91.5
30	26.9	7.0	73.1	93.0	51.4	97.2	91.2

TPH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	32.2	100.0	67.8	23.9	100.0	70.8
5	3.8	16.7	96.2	83.3	36.8	99.5	84.5
10	7.7	14.7	92.3	85.3	38.7	99.1	85.9
15	11.5	11.6	88.5	88.4	43.4	98.7	88.4
20	19.2	4.3	80.8	95.7	65.6	98.0	94.4
25	23.1	2.3	76.9	97.7	76.9	97.7	95.8
30	26.9	2.3	73.1	97.7	76.0	97.3	95.4

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	32.2	100.0	67.8	23.9	100.0	70.8
5	3.8	16.7	96.2	83.3	36.8	99.5	84.5
10	7.7	14.7	92.3	85.3	38.7	99.1	85.9
15	11.5	11.6	88.5	88.4	43.4	98.7	88.4
20	19.2	4.3	80.8	95.7	65.6	98.0	94.4
25	23.1	2.3	76.9	97.7	76.9	97.7	95.8
30	26.9	2.3	73.1	97.7	76.0	97.3	95.4

c. HY28M Small Data Set SQS/SL1

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	58.7	100.0	41.3	18.2	100.0	48.1
5	3.7	40.3	96.3	59.7	23.9	99.2	63.9
10	7.4	30.6	92.6	69.4	28.4	98.6	72.1
15	14.8	24.3	85.2	75.7	31.5	97.5	76.8
20	18.5	16.5	81.5	83.5	39.3	97.2	83.3
25	22.2	11.7	77.8	88.3	46.7	96.8	87.1
30	29.6	9.2	70.4	90.8	50.0	95.9	88.4

TPH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	56.8	100.0	43.2	18.8	100.0	49.8
5	3.7	36.4	96.3	63.6	25.7	99.2	67.4
10	7.4	25.2	92.6	74.8	32.5	98.7	76.8
15	14.8	15.0	85.2	85.0	42.6	97.8	85.0
20	18.5	10.7	81.5	89.3	50.0	97.4	88.4
25	22.2	7.3	77.8	92.7	58.3	97.0	91.0
30	29.6	1.9	70.4	98.1	82.6	96.2	94.8

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	56.3	100.0	43.7	18.9	100.0	50.2
5	3.7	35.4	96.3	64.6	26.3	99.3	68.2
10	7.4	23.8	92.6	76.2	33.8	98.7	78.1
15	14.8	12.6	85.2	87.4	46.9	97.8	87.1
20	18.5	7.8	81.5	92.2	57.9	97.4	91.0
25	22.2	4.9	77.8	95.1	67.7	97.0	93.1
30	29.6	1.9	70.4	98.1	82.6	96.2	94.8

d. HY28M Small Data Set CSL/SL2

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	27.7	100.0	72.3	25.3	100.0	74.7
5	5.0	12.2	95.0	87.8	42.2	99.5	88.4
10	10.0	8.5	90.0	91.5	50.0	99.0	91.4
15	15.0	4.2	85.0	95.8	65.4	98.6	94.8
20	20.0	2.8	80.0	97.2	72.7	98.1	95.7
25	25.0	1.9	75.0	98.1	78.9	97.7	96.1
30	30.0	1.4	70.0	98.6	82.4	97.2	96.1

TPH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	24.9	100.0	75.1	27.4	100.0	77.3
5	5.0	7.5	95.0	92.5	54.3	99.5	92.7
10	10.0	2.3	90.0	97.7	78.3	99.0	97.0
15	15.0	1.4	85.0	98.6	85.0	98.6	97.4
20	20.0	0.9	80.0	99.1	88.9	98.1	97.4
25	25.0	0.5	75.0	99.5	93.8	97.7	97.4
30	30.0	0.5	70.0	99.5	93.3	97.2	97.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	24.9	100.0	75.1	27.4	100.0	77.3
5	5.0	7.5	95.0	92.5	54.3	99.5	92.7
10	10.0	2.3	90.0	97.7	78.3	99.0	97.0
15	15.0	1.4	85.0	98.6	85.0	98.6	97.4
20	20.0	0.9	80.0	99.1	88.9	98.1	97.4
25	25.0	0.5	75.0	99.5	93.8	97.7	97.4
30	30.0	0.5	70.0	99.5	93.3	97.2	97.0

e. CH10G Large Data Set SQS/SL1

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	58.7	100.0	41.3	19.4	100.0	48.6
5	4.6	47.8	95.4	52.2	22.0	98.8	57.5
10	9.2	41.7	90.8	58.3	23.5	97.8	62.3
15	13.8	40.2	86.2	59.8	23.2	96.8	63.0
20	20.0	35.9	80.0	64.1	24.0	95.8	66.1
25	24.6	33.9	75.4	66.1	23.9	95.0	67.2
30	29.2	27.6	70.8	72.4	26.6	94.6	72.2

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	58.0	100.0	42.0	19.6	100.0	49.1
5	4.6	47.6	95.4	52.4	22.1	98.8	57.7
10	9.2	41.3	90.8	58.7	23.7	97.8	62.7
15	13.8	39.8	86.2	60.2	23.4	96.9	63.4
20	20.0	33.5	80.0	66.5	25.2	95.9	68.2
25	24.6	29.6	75.4	70.4	26.5	95.3	71.0
30	29.2	24.3	70.8	75.7	29.1	94.8	75.0

f. CH10G Large Data Set CSL/SL2

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	54.8	100.0	45.2	15.8	100.0	50.3
5	4.1	43.5	95.9	56.5	18.5	99.3	60.2
10	8.2	37.0	91.8	63.0	20.4	98.7	65.7
15	14.3	36.8	85.7	63.2	19.4	97.7	65.3
20	18.4	30.9	81.6	69.1	21.4	97.3	70.3
25	24.5	25.6	75.5	74.4	23.3	96.7	74.5
30	28.6	20.8	71.4	79.2	26.1	96.4	78.5

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	54.8	100.0	45.2	15.8	100.0	50.3
5	4.1	43.5	95.9	56.5	18.5	99.3	60.2
10	8.2	37.0	91.8	63.0	20.4	98.7	65.7
15	14.3	36.8	85.7	63.2	19.4	97.7	65.3
20	18.4	30.9	81.6	69.1	21.4	97.3	70.3
25	24.5	23.3	75.5	76.7	25.0	96.8	76.6
30	28.6	18.3	71.4	81.7	28.7	96.5	80.8

g. HY28M Large Data Set SQS/SL1

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	59.2	100.0	40.8	23.0	100.0	49.7
5	4.3	53.2	95.7	46.8	24.2	98.4	54.2
10	8.5	40.8	91.5	59.2	28.5	97.5	64.1
15	14.9	26.0	85.1	74.0	36.7	96.6	75.6
20	19.1	21.1	80.9	78.9	40.4	95.9	79.2
25	23.4	13.2	76.6	86.8	50.7	95.4	85.3
30	29.8	6.8	70.2	93.2	64.7	94.6	89.7

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	50.9	100.0	49.1	25.8	100.0	56.7
5	4.3	38.5	95.7	61.5	30.6	98.8	66.7
10	8.5	26.8	91.5	73.2	37.7	98.0	76.0
15	14.9	14.3	85.1	85.7	51.3	97.0	85.6
20	19.1	10.6	80.9	89.4	57.6	96.3	88.1
25	23.4	9.8	76.6	90.2	58.1	95.6	88.1
30	29.8	6.0	70.2	94.0	67.3	94.7	90.4

h. HY28M Large Data Set CSL/SL2

PAH

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	33.3	100.0	66.7	22.1	100.0	69.6
5	3.7	13.7	96.3	86.3	40.0	99.6	87.2
10	7.4	10.9	92.6	89.1	44.6	99.2	89.4
15	14.8	7.7	85.2	92.3	51.1	98.5	91.7
20	18.5	4.9	81.5	95.1	61.1	98.2	93.9
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

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	30.2	100.0	69.8	23.9	100.0	72.4
5	3.7	9.8	96.3	90.2	48.1	99.6	90.7
10	7.4	6.0	92.6	94.0	59.5	99.3	93.9
15	14.8	3.5	85.2	96.5	69.7	98.6	95.5
20	18.5	2.8	81.5	97.2	73.3	98.2	95.8
25	22.2	2.1	77.8	97.9	77.8	97.9	96.2
30	29.6	1.8	70.4	98.2	79.2	97.2	95.8

D.2. Regional Differences

The model was run for the entire data set, as well as separately for the data east of the Cascades and west of the Cascades, although for HY10M there was not enough data east of the Cascades to calculate reliability. 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. The results are shown in Table D-2.

Overall, there were no consistent patterns among the results. For some endpoints/effects levels, west side data were more reliable than east side data, and vice versa. In many cases, patterns in the east side results suggested that there were too few data to conduct an effective reliability assessment, or that one survey was dominating the results. In many, but not all, of the cases the combined data were similar to or slightly better than the west side results alone.

Because no clear patterns could be discerned and the east-side database appears insufficient to stand alone at this time, the entire data set was combined and used to calculate state-wide SQVs. It may be possible in the future to develop regional SQVs once more data have been collected in a wider variety of east-sideareas.

Table D-2. Georegion Comparisons

a. CH10G SQS/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	53.3	100.0	46.7	22.0	100.0	53.7
5	4.8	41.2	95.2	58.8	25.8	98.8	63.6
10	9.7	40.9	90.3	59.1	24.9	97.6	63.2
15	14.5	33.7	85.5	66.3	27.6	96.8	68.8
20	19.4	28.3	80.6	71.7	29.9	96.1	72.8
25	24.2	25.4	75.8	74.6	30.9	95.4	74.7
30	29.0	21.1	71.0	78.9	33.6	94.8	77.9

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	34.0	100.0	66.0	15.8	100.0	68.0
5	0.0	34.0	100.0	66.0	15.8	100.0	68.0
10	0.0	34.0	100.0	66.0	15.8	100.0	68.0
15	0.0	34.0	100.0	66.0	15.8	100.0	68.0
20	0.0	34.0	100.0	66.0	15.8	100.0	68.0
25	0.0	34.0	100.0	66.0	15.8	100.0	68.0
30	0.0	34.0	100.0	66.0	15.8	100.0	68.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	58.7	100.0	41.3	19.4	100.0	48.6
5	4.6	47.8	95.4	52.2	22.0	98.8	57.5
10	9.2	41.7	90.8	58.3	23.5	97.8	62.3
15	13.8	40.2	86.2	59.8	23.2	96.8	63.0
20	20.0	35.9	80.0	64.1	24.0	95.8	66.1
25	24.6	33.9	75.4	66.1	23.9	95.0	67.2
30	29.2	27.6	70.8	72.4	26.6	94.6	72.2

b. CH10G CSL/SL2

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	53.0	100.0	47.0	17.2	100.0	52.2
5	4.3	40.7	95.7	59.3	20.5	99.2	62.9
10	8.5	40.4	91.5	59.6	19.9	98.5	62.7
15	14.9	32.2	85.1	67.8	22.5	97.6	69.5
20	19.1	28.5	80.9	71.5	23.8	97.1	72.4
25	23.4	24.1	76.6	75.9	25.9	96.7	76.0
30	29.8	16.1	70.2	83.9	32.4	96.2	82.5

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	35.4	100.0	64.6	10.5	100.0	66.0
5	0.0	35.4	100.0	64.6	10.5	100.0	66.0
10	0.0	35.4	100.0	64.6	10.5	100.0	66.0
15	0.0	35.4	100.0	64.6	10.5	100.0	66.0
20	0.0	35.4	100.0	64.6	10.5	100.0	66.0
25	0.0	35.4	100.0	64.6	10.5	100.0	66.0
30	0.0	35.4	100.0	64.6	10.5	100.0	66.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	54.8	100.0	45.2	15.8	100.0	50.3
5	4.1	43.5	95.9	56.5	18.5	99.3	60.2
10	8.2	37.0	91.8	63.0	20.4	98.7	65.7
15	14.3	36.8	85.7	63.2	19.4	97.7	65.3
20	18.4	30.9	81.6	69.1	21.4	97.3	70.3
25	24.5	25.6	75.5	74.4	23.3	96.7	74.5
30	28.6	20.8	71.4	79.2	26.1	96.4	78.5

c. CH10M SQS/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	96.1	100.0	3.9	27.2	100.0	29.3
5	4.4	81.9	95.6	18.1	29.6	92.0	38.6
10	9.5	74.0	90.5	26.0	30.5	88.4	43.1
15	14.6	67.2	85.4	32.8	31.4	86.2	46.7
20	19.7	61.2	80.3	38.8	32.1	84.6	49.8
25	24.8	56.7	75.2	43.3	32.3	82.9	51.7
30	29.9	48.3	70.1	51.7	34.3	82.8	56.6

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	76.5	100.0	23.5	38.1	100.0	48.0
5	0.0	76.5	100.0	23.5	38.1	100.0	48.0
10	6.3	58.8	93.8	41.2	42.9	93.3	58.0
15	6.3	58.8	93.8	41.2	42.9	93.3	58.0
20	18.8	32.4	81.3	67.6	54.2	88.5	72.0
25	18.8	32.4	81.3	67.6	54.2	88.5	72.0
30	18.8	32.4	81.3	67.6	54.2	88.5	72.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	97.3	100.0	2.7	27.5	100.0	28.9
5	4.6	83.4	95.4	16.6	29.7	90.8	37.9
10	9.8	71.8	90.2	28.2	31.7	88.6	44.9
15	14.4	63.4	85.6	36.6	33.2	87.4	49.8
20	19.6	52.0	80.4	48.0	36.3	86.9	56.7
25	24.8	46.0	75.2	54.0	37.6	85.5	59.7
30	29.4	40.7	70.6	59.3	39.0	84.5	62.3

d. CH10M CSL/SL2

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	52.8	100.0	47.2	19.9	100.0	53.3
5	5.0	42.8	95.0	57.2	22.5	98.9	61.6
10	10.0	33.0	90.0	67.0	26.3	98.1	69.7
15	15.0	23.6	85.0	76.4	32.1	97.5	77.4
20	20.0	19.4	80.0	80.6	35.0	96.9	80.5
25	25.0	17.7	75.0	82.3	35.7	96.2	81.5
30	30.0	13.1	70.0	86.9	41.2	95.7	84.9

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	25.6	100.0	74.4	38.9	100.0	78.0
5	0.0	25.6	100.0	74.4	38.9	100.0	78.0
10	0.0	25.6	100.0	74.4	38.9	100.0	78.0
15	14.3	18.6	85.7	81.4	42.9	97.2	82.0
20	14.3	18.6	85.7	81.4	42.9	97.2	82.0
25	14.3	18.6	85.7	81.4	42.9	97.2	82.0
30	28.6	11.6	71.4	88.4	50.0	95.0	86.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	50.9	100.0	49.1	20.8	100.0	55.1
5	4.5	42.7	95.5	57.3	23.0	99.0	61.8
10	9.0	37.1	91.0	62.9	24.7	98.1	66.2
15	14.9	29.1	85.1	70.9	28.1	97.3	72.5
20	19.4	20.8	80.6	79.2	34.2	96.8	79.4
25	23.9	18.0	76.1	82.0	36.2	96.3	81.3
30	29.9	13.0	70.1	87.0	42.0	95.6	85.0

e. HY10M SQS/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	93.7	100.0	6.3	30.6	100.0	33.7
5	4.8	78.7	95.2	21.3	33.3	91.5	42.9
10	9.5	71.7	90.5	28.3	34.3	87.8	46.5
15	14.3	65.0	85.7	35.0	35.3	85.6	49.9
20	20.0	54.7	80.0	45.3	37.7	84.6	55.4
25	24.8	40.2	75.2	59.8	43.6	85.4	64.3
30	29.5	36.2	70.5	63.8	44.6	83.9	65.7

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	95.7	100.0	4.3	30.7	100.0	32.8
5	4.6	82.9	95.4	17.1	32.8	89.8	40.4
10	9.2	72.4	90.8	27.6	34.7	87.7	46.4
15	14.7	60.7	85.3	39.3	37.3	86.3	53.0
20	19.3	49.8	80.7	50.2	40.7	86.0	59.3
25	24.8	46.7	75.2	53.3	40.6	83.5	59.8
30	29.4	44.7	70.6	55.3	40.1	81.6	59.8

f. HY10M CSL/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	59.4	100.0	40.6	21.0	100.0	48.7
5	4.1	57.4	95.9	42.6	20.9	98.5	49.9
10	8.2	55.2	91.8	44.8	20.8	97.2	51.3
15	14.3	48.7	85.7	51.3	21.8	95.8	56.0
20	18.4	44.5	81.6	55.5	22.5	95.0	59.1
25	24.5	30.0	75.5	70.0	28.5	94.8	70.8
30	28.6	23.2	71.4	76.8	32.7	94.4	76.0

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	69.1	100.0	30.9	19.3	100.0	40.7
5	3.8	54.5	96.2	45.5	22.6	98.6	52.7
10	9.6	49.0	90.4	51.0	23.4	97.0	56.6
15	13.5	35.4	86.5	64.6	28.8	96.7	67.8
20	19.2	32.2	80.8	67.8	29.4	95.5	69.7
25	25.0	25.8	75.0	74.2	32.5	94.7	74.3
30	28.8	24.5	71.2	75.5	32.5	94.0	74.9

g. HY28G SQS/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	83.1	100.0	16.9	33.2	100.0	41.2
5	4.1	74.0	95.9	26.0	34.8	93.9	46.4
10	9.6	62.1	90.4	37.9	37.5	90.5	53.2
15	13.7	58.8	86.3	41.2	37.7	88.0	54.4
20	19.2	48.6	80.8	51.4	40.7	86.7	60.0
25	24.7	43.5	75.3	56.5	41.7	84.7	62.0
30	28.8	42.9	71.2	57.1	40.6	82.8	61.2

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	54.5	100.0	45.5	36.8	100.0	58.6
5	0.0	54.5	100.0	45.5	36.8	100.0	58.6
10	7.1	27.3	92.9	72.7	52.0	97.0	77.6
15	14.3	25.0	85.7	75.0	52.2	94.3	77.6
20	14.3	25.0	85.7	75.0	52.2	94.3	77.6
25	21.4	15.9	78.6	84.1	61.1	92.5	82.8
30	28.6	11.4	71.4	88.6	66.7	90.7	84.5

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	79.6	100.0	20.4	33.1	100.0	42.9
5	4.6	74.7	95.4	25.3	33.5	93.3	45.1
10	9.2	70.6	90.8	29.4	33.6	89.0	46.8
15	14.9	63.8	85.1	36.2	34.4	86.0	50.0
20	19.5	62.0	80.5	38.0	33.8	83.2	50.0
25	24.1	58.4	75.9	41.6	33.8	81.4	51.3
30	29.9	57.0	70.1	43.0	32.6	78.5	50.6

h. HY28G CSL/SL2

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	28.6	100.0	71.4	11.5	100.0	72.4
5	0.0	28.6	100.0	71.4	11.5	100.0	72.4
10	0.0	28.6	100.0	71.4	11.5	100.0	72.4
15	11.1	28.6	88.9	71.4	10.4	99.4	72.0
20	11.1	28.6	88.9	71.4	10.4	99.4	72.0
25	22.2	10.8	77.8	89.2	21.2	99.1	88.8
30	22.2	10.8	77.8	89.2	21.2	99.1	88.8

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	12.2	100.0	87.8	60.0	100.0	89.7
5	0.0	12.2	100.0	87.8	60.0	100.0	89.7
10	0.0	12.2	100.0	87.8	60.0	100.0	89.7
15	11.1	4.1	88.9	95.9	80.0	97.9	94.8
20	11.1	4.1	88.9	95.9	80.0	97.9	94.8
25	22.2	4.1	77.8	95.9	77.8	95.9	93.1
30	22.2	4.1	77.8	95.9	77.8	95.9	93.1

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	36.2	100.0	63.8	14.6	100.0	65.9
5	0.0	36.2	100.0	63.8	14.6	100.0	65.9
10	5.6	36.2	94.4	63.8	13.9	99.5	65.6
15	11.1	23.8	88.9	76.2	18.8	99.1	76.9
20	16.7	17.9	83.3	82.1	22.4	98.8	82.1
25	22.2	9.7	77.8	90.3	33.3	98.5	89.6
30	27.8	6.6	72.2	93.4	40.6	98.2	92.2

i. HY28M SQS/SL1

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	64.2	100.0	35.8	20.5	100.0	44.9
5	2.8	53.2	97.2	46.8	23.2	99.0	53.9
10	8.3	37.2	91.7	62.8	28.9	97.9	66.9
15	13.9	27.1	86.1	72.9	34.4	97.0	74.8
20	19.4	20.2	80.6	79.8	39.7	96.1	79.9
25	25.0	14.2	75.0	85.8	46.6	95.4	84.3
30	27.8	10.1	72.2	89.9	54.2	95.1	87.4

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	25.5	100.0	74.5	47.8	100.0	79.3
5	0.0	25.5	100.0	74.5	47.8	100.0	79.3
10	9.1	25.5	90.9	74.5	45.5	97.2	77.6
15	9.1	25.5	90.9	74.5	45.5	97.2	77.6
20	18.2	21.3	81.8	78.7	47.4	94.9	79.3
25	18.2	21.3	81.8	78.7	47.4	94.9	79.3
30	27.3	21.3	72.7	78.7	44.4	92.5	77.6

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	59.2	100.0	40.8	23.0	100.0	49.7
5	4.3	53.2	95.7	46.8	24.2	98.4	54.2
10	8.5	40.8	91.5	59.2	28.5	97.5	64.1
15	14.9	26.0	85.1	74.0	36.7	96.6	75.6
20	19.1	21.1	80.9	78.9	40.4	95.9	79.2
25	23.4	13.2	76.6	86.8	50.7	95.4	85.3
30	29.8	6.8	70.2	93.2	64.7	94.6	89.7

j. HY28M CSL/SL2

West

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	29.9	100.0	70.1	25.0	100.0	72.8
5	4.3	13.9	95.7	86.1	40.7	99.5	87.0
10	8.7	9.5	91.3	90.5	48.8	99.1	90.6
15	13.0	6.5	87.0	93.5	57.1	98.6	92.9
20	17.4	4.8	82.6	95.2	63.3	98.2	94.1
25	21.7	3.5	78.3	96.5	69.2	97.8	94.9
30	26.1	2.2	73.9	97.8	77.3	97.4	95.7

East

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	5.6	100.0	94.4	57.1	100.0	94.8
5	0.0	5.6	100.0	94.4	57.1	100.0	94.8
10	0.0	5.6	100.0	94.4	57.1	100.0	94.8
15	0.0	5.6	100.0	94.4	57.1	100.0	94.8
20	0.0	5.6	100.0	94.4	57.1	100.0	94.8
25	25.0	5.6	75.0	94.4	50.0	98.1	93.1
30	25.0	5.6	75.0	94.4	50.0	98.1	93.1

Combined

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	33.3	100.0	66.7	22.1	100.0	69.6
5	3.7	13.7	96.3	86.3	40.0	99.6	87.2
10	7.4	10.9	92.6	89.1	44.6	99.2	89.4
15	14.8	7.7	85.2	92.3	51.1	98.5	91.7
20	18.5	4.9	81.5	95.1	61.1	98.2	93.9
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

D.3. Comparison to Reference vs. Control

The subset of the data set that includes reference data was used to evaluate the reliability of comparison to control vs. comparison to reference, to test the previous finding (SAIC and Avocet, 2003) that comparison to control provides similar or better reliability than comparison to reference. The results are shown in Table D-3 below. In these tables, green colored regions are those that performed better in a given table.

In all cases, comparison to control and reference were equally good or comparison to control was much better. Therefore, the workgroup chose to use comparison to control, in part due to these reliability evaluations and in part because there are far more data available to work with.

Table D-3. Comparison to Reference vs. Control

a. CH10G SQS/SL1

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	63.4	100.0	36.6	21.8	100.0	46.1
5	3.7	62.7	96.3	37.3	21.3	98.3	46.1
10	7.4	51.6	92.6	48.4	24.0	97.4	55.0
15	14.8	41.2	85.2	58.8	26.7	95.7	62.8
20	18.5	37.9	81.5	62.1	27.5	95.0	65.0
25	22.2	28.8	77.8	71.2	32.3	94.8	72.2
30	29.6	5.2	70.4	94.8	70.4	94.8	91.1

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	70.8	100.0	29.2	35.2	100.0	48.9
5	4.0	57.7	96.0	42.3	39.0	96.5	57.2
10	10.0	40.8	90.0	59.2	45.9	93.9	67.8
15	14.0	28.5	86.0	71.5	53.8	93.0	75.6
20	20.0	20.0	80.0	80.0	60.6	91.2	80.0
25	24.0	19.2	76.0	80.8	60.3	89.7	79.4
30	30.0	15.4	70.0	84.6	63.6	88.0	80.6

Conclusion: Neither is clearly better.

b. CH10G CSL/SL2

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	32.1	100.0	67.9	29.2	100.0	71.7
5	4.8	14.5	95.2	85.5	46.5	99.3	86.7
10	9.5	9.4	90.5	90.6	55.9	98.6	90.6
15	14.3	8.8	85.7	91.2	56.3	98.0	90.6
20	19.0	6.9	81.0	93.1	60.7	97.4	91.7
25	23.8	3.1	76.2	96.9	76.2	96.9	94.4
30	28.6	1.9	71.4	98.1	83.3	96.3	95.0

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	37.0	100.0	63.0	38.6	100.0	70.0
5	2.9	26.0	97.1	74.0	46.5	99.1	78.3
10	8.8	21.2	91.2	78.8	50.0	97.5	81.1
15	14.7	13.7	85.3	86.3	59.2	96.2	86.1
20	17.6	13.0	82.4	87.0	59.6	95.5	86.1
25	23.5	11.0	76.5	89.0	61.9	94.2	86.7
30	29.4	7.5	70.6	92.5	68.6	93.1	88.3

Conclusion: Control is clearly better.

c. CH10M SQS/SL1

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	72.4	100.0	27.6	43.3	100.0	53.4
5	4.1	59.7	95.9	40.3	47.0	94.7	60.1
10	9.5	53.0	90.5	47.0	48.6	90.0	62.5
15	14.9	47.8	85.1	52.2	49.6	86.4	63.9
20	18.9	43.3	81.1	56.7	50.8	84.4	65.4
25	24.3	35.8	75.7	64.2	53.8	82.7	68.3
30	29.7	26.9	70.3	73.1	59.1	81.7	72.1

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	64.5	100.0	35.5	17.5	100.0	43.3
5	4.0	61.7	96.0	38.3	17.5	98.6	45.2
10	8.0	56.8	92.0	43.2	18.1	97.5	49.0
15	12.0	54.6	88.0	45.4	18.0	96.5	50.5
20	20.0	47.5	80.0	52.5	18.7	95.0	55.8
25	24.0	40.4	76.0	59.6	20.4	94.8	61.5
30	28.0	37.2	72.0	62.8	20.9	94.3	63.9

Conclusion: Control is slightly better.

d. CH10M CSL/SL2

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	42.2	100.0	57.8	26.9	100.0	63.5
5	3.6	41.1	96.4	58.9	26.7	99.1	63.9
10	7.1	40.0	92.9	60.0	26.5	98.2	64.4
15	14.3	36.1	85.7	63.9	27.0	96.6	66.8
20	17.9	31.1	82.1	68.9	29.1	96.1	70.7
25	25.0	26.1	75.0	73.9	30.9	95.0	74.0
30	28.6	21.1	71.4	78.9	34.5	94.7	77.9

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	35.4	100.0	64.6	12.5	100.0	66.3
5	0.0	35.4	100.0	64.6	12.5	100.0	66.3
10	10.0	33.8	90.0	66.2	11.8	99.2	67.3
15	10.0	33.8	90.0	66.2	11.8	99.2	67.3
20	20.0	33.3	80.0	66.7	10.8	98.5	67.3
25	20.0	33.3	80.0	66.7	10.8	98.5	67.3
30	30.0	18.2	70.0	81.8	16.3	98.2	81.3

Conclusion: Neither is clearly better.

e. HY28G SQS/SL1

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	60.7	100.0	39.3	48.5	100.0	61.4
5	0.0	60.7	100.0	39.3	48.5	100.0	61.4
10	6.3	32.1	93.8	67.9	62.5	95.0	77.3
15	12.5	10.7	87.5	89.3	82.4	92.6	88.6
20	18.8	0.0	81.3	100.0	100.0	90.3	93.2
25	18.8	0.0	81.3	100.0	100.0	90.3	93.2
30	18.8	0.0	81.3	100.0	100.0	90.3	93.2

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	50.0	100.0	50.0	62.5	100.0	72.7
5	5.0	45.8	95.0	54.2	63.3	92.9	72.7
10	10.0	45.8	90.0	54.2	62.1	86.7	70.5
15	15.0	45.8	85.0	54.2	60.7	81.3	68.2
20	20.0	41.7	80.0	58.3	61.5	77.8	68.2
25	25.0	25.0	75.0	75.0	71.4	78.3	75.0
30	30.0	16.7	70.0	83.3	77.8	76.9	77.3

Conclusion: Control is clearly better.

f. HY28G CSL/SL2

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	7.7	100.0	92.3	62.5	100.0	93.2
5	0.0	7.7	100.0	92.3	62.5	100.0	93.2
10	0.0	7.7	100.0	92.3	62.5	100.0	93.2
15	0.0	7.7	100.0	92.3	62.5	100.0	93.2
20	20.0	5.1	80.0	94.9	66.7	97.4	93.2
25	20.0	5.1	80.0	94.9	66.7	97.4	93.2
30	20.0	5.1	80.0	94.9	66.7	97.4	93.2

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	40.6	100.0	59.4	48.0	100.0	70.5
5	0.0	40.6	100.0	59.4	48.0	100.0	70.5
10	8.3	34.4	91.7	65.6	50.0	95.5	72.7
15	8.3	34.4	91.7	65.6	50.0	95.5	72.7
20	16.7	15.6	83.3	84.4	66.7	93.1	84.1
25	16.7	15.6	83.3	84.4	66.7	93.1	84.1
30	16.7	15.6	83.3	84.4	66.7	93.1	84.1

Conclusion: Control is clearly better.

g. HY28M SQS/SL1

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	19.4	100.0	80.6	68.4	100.0	86.4
5	0.0	19.4	100.0	80.6	68.4	100.0	86.4
10	7.7	19.4	92.3	80.6	66.7	96.2	84.1
15	7.7	19.4	92.3	80.6	66.7	96.2	84.1
20	15.4	12.9	84.6	87.1	73.3	93.1	86.4
25	23.1	3.2	76.9	96.8	90.9	90.9	90.9
30	23.1	3.2	76.9	96.8	90.9	90.9	90.9

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	21.2	100.0	78.8	61.1	100.0	84.1
5	0.0	21.2	100.0	78.8	61.1	100.0	84.1
10	9.1	21.2	90.9	78.8	58.8	96.3	81.8
15	9.1	21.2	90.9	78.8	58.8	96.3	81.8
20	18.2	15.2	81.8	84.8	64.3	93.3	84.1
25	18.2	15.2	81.8	84.8	64.3	93.3	84.1
30	27.3	6.1	72.7	93.9	80.0	91.2	88.6

Conclusion: Control is clearly better.

h. HY28M CSL/SL2

Control

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	7.3	100.0	92.7	50.0	100.0	93.2
5	0.0	7.3	100.0	92.7	50.0	100.0	93.2
10	0.0	7.3	100.0	92.7	50.0	100.0	93.2
15	0.0	7.3	100.0	92.7	50.0	100.0	93.2
20	0.0	7.3	100.0	92.7	50.0	100.0	93.2
25	0.0	7.3	100.0	92.7	50.0	100.0	93.2
30	0.0	7.3	100.0	92.7	50.0	100.0	93.2

Reference

Nominal Percentiles	%False Negatives	%False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
0	0.0	7.3	100.0	92.7	50.0	100.0	93.2
5	0.0	7.3	100.0	92.7	50.0	100.0	93.2
10	0.0	7.3	100.0	92.7	50.0	100.0	93.2
15	0.0	7.3	100.0	92.7	50.0	100.0	93.2
20	0.0	7.3	100.0	92.7	50.0	100.0	93.2
25	0.0	7.3	100.0	92.7	50.0	100.0	93.2
30	0.0	7.3	100.0	92.7	50.0	100.0	93.2

Conclusion: Both are the same.

D.4. Blank-Correction

It was determined during the quality assurance review that the data sets had not all been blank-corrected in the same manner. Furthermore, a number of chemicals known to be laboratory contaminants and not likely to be found in environmental sediments were associated with false positives in the data set, including acetone (5 false positives) and methylene chloride (57 false positives). This issue was resolved by applying EPA contract laboratory protocol blank-correction methods to all of the data in all of the historical data sets consistently, revising qualifier codes as necessary.

Following this step, the model was begun again to evaluate the effect of this change in the data set. It was immediately apparent that this requalification had improved the results, because the data set no longer contained acetone, methylene chloride, or isopropylbenzene, among other chemicals that are highly unlikely to be found in sediments but are common laboratory contaminants. These chemicals no longer had enough detections to pass the initial screening criteria. In addition, it is likely that some spurious results for chemicals that can be found in the environment but are also common laboratory contaminants were removed, leaving only those detections more likely to be associated with actual environmental concentrations.

Because this evaluation was conducted by examining the data set itself and the initial data screening results, reliability analysis was not conducted for this step alone. However, this process along with a number of other more minor quality assurance screening evaluations of the data significantly improved reliability in incremental steps.