DEVELOPMENT OF FRESHWATER SEDIMENT QUALITY VALUES FOR USE IN WASHINGTON STATE

PHASE I TASK 6: FINAL REPORT

SEPTEMBER 2002

Publication Number 02-09-050

prepared for:

Washington Department of Ecology
Sediment Management Unit

prepared by:

Science Applications International Corporation

Avocet Consulting

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

List of Acronyms

AEL - Average Effects Level
AET – Apparent Effects Threshold
ASTM – American Society for Testing and Materials.
AVS – Acid-volatile Sulfides
BSA – Bioassay Statistical Analysis Tool
BTEX – bis(2-ethylhexyl)phthalate
CF/CI – Control Final Divided by Control Initial Value
COE – U.S. Army Corps of Engineers, Portland District
CSL – Cleanup Screening Level
DEQ – Department of Environmental Quality (Oregon)
EqP – Equilibrium Partitioning
ERLs – Effects Range Low
ERMs – Effects Range Medium
FSQV – Freshwater Sediment Quality Value
GLNPO – Great Lakes National Program Office
HPAH – High Molecular Weight Polynuclear Aromatic Hydrocarbon
LAET – Lowest Apparent Effects Thresholds
LEL – Lowest Effect Level
LPAET – Lowest Probable Apparent Effects Thresholds
LPAH – Low Molecular Weight Polynuclear Aromatic Hydrocarbon
MET – Minimum Effects Threshold
NEC – No Effects Concentration
PAETs – Probable Apparent Effects Thresholds
PCBs – Polychlorinated Biphenyls
PEC – Probable Effects Concentration
PEL – Probable Effects Levels
PSDDA – Puget Sound Dredged Disposal Analysis
QA – Quality Assurance
SEDQUAL – Washington State Department of Ecology's Sediment Quality Database
SEL – Severe Effect Level
SEM – Simultaneously-extracted Metals
SLC – Screening Level Concentration
SMS – Sediment Management Standards (Washington)
SQS – Sediment Quality Standard
SQV – Sediment Quality Value
TEC – Threshold Effects Concentration
TEL – Threshold Effects Levels
TET – Toxic Effects Threshold
TMDL – Total Maximum Daily Load
TOC – Total Organic Carbon
USGS – U.S. Geological Society

Qualifier Definitions

U qualifier – Undetected at the detection limit shown
B qualifier – Detected in samples and associated method blank - PSEP
X qualifier – Recovery less than 10%

Executive Summary

Report Objective

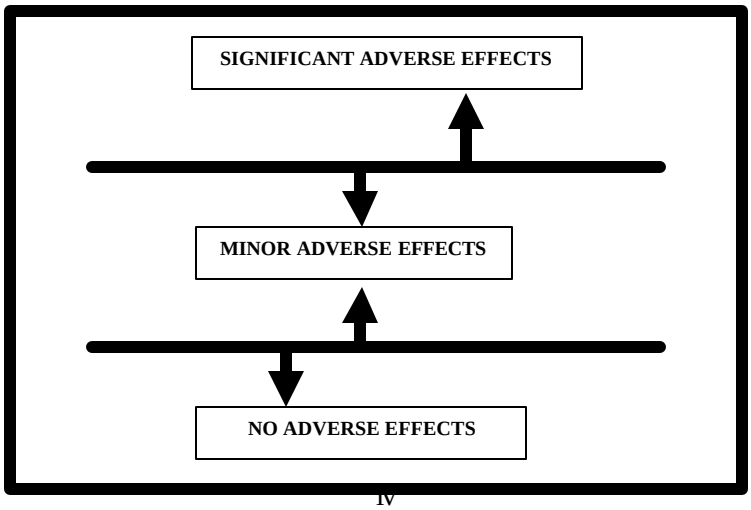
This report describes Ecology's most recent effort in the evaluation of freshwater sediment quality values (SQVs) for possible use in Ecology's sediment management program. The present effort includes a compilation, description, and evaluation of existing freshwater SQV sets in North America. A freshwater SQV set is a compilation of values for multiple, individual chemical parameters. A sediment quality exceedances of any individual chemical parameter in the SQV set would predict that adverse biological effects will occur in freshwater sediments. These evaluations are the first step toward identifying, developing and using freshwater SQVs for regulatory decision-making, e.g., contaminated freshwater sediment cleanup site identification, in Washington State.

The present analysis addresses the development of a mutli-chemical SQV set for general application state-wide. It is not intended to supersede action or cleanup levels that have been or will be developed for individual chemicals at specific sites. In addition, the present analysis considers only ecological effects, not human health effects; SQVs based on ecological effects could be different from those based on human health effects.

Methods

Approximately eighteen North American SQV sets were identified and from these a subset of eight SQV sets were prioritized for evaluation. Each SQV set was evaluated using a regional freshwater sediment laboratory bioassay data-set assembled and entered into the SEDQUAL Information System. The evaluation, termed a reliability assessment, identified the ability of each SQV set to correctly predict biological effects and to correctly predict biological non-effects.

This evaluation centered around the Sediment Management Standards, Chapter 173-204 WAC, two-parallel line paradigm for regulatory application of sediment quality values. The lower of these lines currently represents the federally approved sediment quality standard and no adverse effects level, above which sediment quality regulatory assessment (which could include additional testing and evaluation for source control, cleanup and dredged material disposal evaluations) is necessary. The upper parallel line represents the division between minor and significant adverse effects used to establish sediment quality regulatory limitations (see below). For example, currently the SMS rule uses exceedances of the upper line to trigger cleanup site identification.



Results

Ecology's evaluation focused on identification of a SQV set that acceptably minimized both false positive and false negative predictions. Ideally, one SQV set could be used to screen out clean sediments and screen in contaminated sediments requiring further regulatory assessment. The evaluation found:

- No one SQV set can currently meet both screening and regulatory evaluation needs. In general, the evaluation found that the conservative SQV sets (low/stringent values) suffered from high rates of predicted false positives, (calling a biological non-hit a hit). Conversely, the less-stringent SQV sets (high values) suffered from high rates of predicted false negatives, (calling a biological hit a non-hit).

- The evaluation did identify individual SQV sets that could be combined to de-prioritize freshwater sediments below the no adverse effects line and prioritize sediments that require further regulatory analyses, i.e., above the upper line. It is likely that the majority of contaminated freshwater sediments will fall in-between these lines; and

- To support the evaluation, Ecology completed a comprehensive re-development of the SEDQUAL Bioassay Statistical Analysis tool. These changes added the ability to evaluate freshwater sediment bioassay performance and allow default or user-stipulated performance criteria.

Recommendations

In most cases, it is still recommended that the use of chemistry **and** biological testing be used to evaluate freshwater sediment quality for the appropriate regulatory response. Biological testing via laboratory bioassays and/or assessment of benthic assemblages is recommended for most freshwater sediment regulatory assessments for the immediate future.

Finally, the results of this study identified limited Phase II future work recommendations centered around calculation of new sediment quality values based on the Apparent Effects Threshold and Floating Percentile methods. Use of these new SQV sets would be tested via the methods noted above to identify whether improved predictive reliability will allow use of a SQV set alone for freshwater sediment quality assessments.

Report Organization

Section 1.0 of this report describes the existing SQV sets currently in use in North America and identifies those most relevant to sediment management in Washington State (which are evaluated further in this study). Section 2.0 describes the collection and screening of additional freshwater sediment chemical and toxicological data from the Northwest that were added to the SEDQUAL database for use in the present analysis. Section 3.0 describes the results of the reliability analysis of the existing SQV sets, which evaluated the ability of these SQV sets to correctly predict sediment toxicity. The Appendices contain data, calculations and other information in support of each of the major sections of this report.

1.0 SUMMARY OF EXISTING SEDIMENT QUALITY VALUES IN NORTH AMERICA (TASK 2)

1.1 Introduction

This section provides a compilation, description, and evaluation of existing freshwater sediment quality value (SQV) sets in North America, as the first step toward identifying or developing freshwater SQVs for use in Washington State. This section provides the results of Task 2, the goals of which were to:

- Provide an up-to-date compilation of freshwater SQV sets in North America
- Identify up to five SQV sets to be carried forward for reliability testing in Task 5 (Section 3.0)
- Provide technical and regulatory information on each SQV set which, along with the results of reliability testing from Task 5, form the basis of recommendations to Ecology for use of the SQV sets in Ecology's freshwater sediment management programs. Final guidance on the use of these SQVs will be prepared by Ecology.

The SQVs were compiled by contacting agency staff in U.S. states and Canadian provinces known to be active in regulating sediments; sediment quality experts from national, state, and provincial agencies; and academic experts and consultants active in the field. Contact information is provided in the technical appendices (Appendix I).

Section 1.2 provides a summary overview of each set of freshwater SQVs that was identified, including information on the scientific approach used to calculate the SQVs, their geographic scope, the biological tests used to calculate effects-based SQVs, basic information on units, normalization, and summing of the numeric SQVs, the narrative or policy basis for the SQVs, and their current use in regulatory programs.

Section 1.3 provides a comparison of the SQV sets with each other and against a set of criteria intended to evaluate the SQV sets according to their suitability for use within Ecology's freshwater sediment management programs and to carry forward for reliability testing as part of Task 5. The reliability testing was intended to determine whether any of these existing SQV sets adequately predicts the presence and absence of biological effects in Pacific Northwest freshwater sediments. A summary table of the rankings is provided, and the numeric SQVs are listed in Appendix H.

> *Note: These SQVs have been compiled for reference and evaluation purposes only. Their presence in this report and the publication of this information in no way implies that any of the numeric SQVs or methods contained herein are currently endorsed or recommended by the Department of Ecology. The evaluations conducted in this report and subsequent tasks are for the sole purpose of providing contractor recommendations to Ecology. Ecology will develop final guidance on how the SQV sets should be used in its regulatory programs.*

Section 1.6 discusses a number of potential approaches that could be used to derive SQVs based on freshwater benthic data. These approaches are in use in various regions of the US and Canada to evaluate the health of freshwater ecosystems, but have not yet been used in conjunction with chemistry data to derive SQVs. These benthic community approaches are included because of their scientific and regulatory importance as a component of the chemistry-bioassay-benthic community triad and as an indicator of potential chronic effects. They represent possible approaches that could be considered as part of Phase II to ensure that this component is incorporated into Washington State's freshwater SQVs.

Section 1.7 provides a summary and conclusions for Task 2.

The technical appendices provide more detailed information on each SQV set described in Section 1.2, including:

- identification of authors and contact information
- bibliographic references
- a description of the data sets used to calculate the SQVs
- electronic data that may be available
- a detailed description of calculation methods
- a description of the quality assurance procedures used
- information on any reliability assessments that may already have been performed

In addition, a summary of the numeric SQVs associated with these different approaches is provided in Appendix H.

1.2 Sediment Quality Value Summaries

Each of the SQV sets that were compiled are described in summary format. The information provided below is primarily derived from agency documents and websites, and represents the information provided by the agency that developed the SQVs. Particularly with respect to regulatory use, it is not always possible to determine how these values are used in actual practice. Additional details on each set of SQVs are provided in the technical appendices for each set of SQVs. These summary descriptions, along with some of the material in the appendices, are used in the comparison and evaluation in Section 1.3.

The following sets of freshwater SQVs were identified and are discussed below:

- Apparent Effects Thresholds and Probable Apparent Effects Thresholds developed by the Washington Department of Ecology
- No-Effects Concentrations developed by USGS/GLNPO
- Effects Range Low and Effects Range Median, developed by USGS/GLNPO
- Threshold Effects Levels and Probable Effects Levels developed by USGS/GLNPO
- Threshold Effects Levels and Probable Effects Levels developed by MacDonald et al. for the Canadian Council of Ministers of the Environment
- Average Effects Levels and Probable Effects Levels developed by MacDonald for British Columbia
- Consensus-Based Sediment Quality Guidelines developed by Ingersoll et al. for the St. Louis Area of Concern
- Screening Level Concentrations developed by the Ontario Ministry of the Environment
- Screening Level Concentrations developed by Quebec for the St. Lawrence River
- Equilibrium Partitioning Values developed by the US Environmental Protection Agency
- Equilibrium Partitioning Values developed by New York Department of Conservation

In several cases, the discussions of these SQVs are combined into one section where they are sufficiently similar.

1.2.1 APPARENT EFFECTS THRESHOLDS AND PROBABLE APPARENT EFFECTS THRESHOLDS

Overall Approach. In 1997, the Washington Department of Ecology published a set of freshwater Apparent Effects Thresholds (AETs) and Probable AETs (PAETs), which were not formalized as guidelines or criteria. AETs are levels above which adverse effects have always been observed in biological tests in the data set used to calculate them. First, outliers are removed from the data set (unusually high no-effects data), and then the highest remaining no-effect value sets the AET. Due to concerns about outlier analysis methods, a modification to this approach was proposed termed the Probable AET, in which the 95th percentile of the no-effects distribution (including only those values higher than the lowest effects concentration) is selected as the PAET. Additional details on calculation methods are provided in Appendix A.

Normalization and Units. AETs and PAETs were calculated for 15 individual PAHs, as well as LPAH and HPAH sums, 5 miscellaneous organics such as phthalates and phenol, 5 chlorinated organics including dioxins and PCBs, 12 metals, sulfides, ammonia, and total organic carbon (TOC). AETs were calculated using two different normalization methods – one with all AETs in dry weight, and the other using a mixed approach in which AETs are normalized to organic carbon for non-polar organic compounds and to dry weight for polar organics, metals, and conventionals. AETs are expressed in mg/kg for metals, nonpolar organic-carbon normalized organic chemicals, and ammonia/sulfides; in µg/kg for polar organics, nonpolar organics normalized to dry weight, and tributyltin; and in percent for TOC.

Geographic Scope. In developing the freshwater AETs, Ecology used 34 surveys with 245 stations from freshwater areas in Washington and the lower Willamette River in Oregon, most of which were from the lower Columbia River or freshwater lakes west of the Cascade Mountains. A few older surveys from eastern Washington (upper Columbia River) were also included. At the time, this was all the synoptic chemistry and bioassay data known to exist in the region.

Biological Tests Included. In theory, AETs would be calculated for several different biological tests, then arranged in order of increasing concentration. The lowest AET is designed to correspond to a level below which adverse effects would not be expected. Between the lowest AET and the second-lowest AET, some minor adverse effects may occur (but biological testing is recommended to confirm the presence or absence of effects). Above the second-lowest AET, more significant adverse effects could be expected and/or effects may occur in a greater percentage of benthic species. This level is considered unacceptable in Ecology's regulatory programs, unless biological testing determines that adverse effects are not occurring to the degree predicted. At the time these AETs were published, however, there was not enough data for a variety of acute and chronic biological tests to develop the full range of AETs that would support this regulatory approach.

Of the biological tests conducted at the included stations, the vast majority (228) were for the acute bioassay using *Hyalella azteca*. Also included were 60 Microtox stations, and lesser numbers of *Ceriodaphnia, Daphnia, Chironomus,* and *Hexagenia* stations (several surveys had more than one bioassay at each station). Of these tests, only *Hyalella* and Microtox were considered to have sufficient data to calculate AETs. Although some benthic community surveys were identified, they were not entered into the database and were not considered in this effort.

Regulatory Use. Freshwater AETs are not yet in routine use in regulatory programs, although they are among the existing SQVs that are considered for use in Washington State on a case-by-case basis, and have also been evaluated for use in Portland Harbor, Oregon. Marine AETs, in contrast, have been available since 1986 for a variety of acute and chronic biological tests, including benthic community

measures. Marine AETs were updated in 1988, and again in 1998 and 1999, to include additional data and biological tests.

AETs have been used successfully in interagency Puget Sound dredging programs since 1988 to determine the suitability of dredged material for open-water disposal. AETs were adopted by regulation as marine sediment quality standards for Ecology's cleanup and source control programs in 1991, and were subsequently approved by EPA Region 10 as federally recognized water quality criteria for the State of Washington. The marine AETs were also included in 1998 as part of the Dredged Material Evaluation Framework for the Lower Columbia River dredging program. Monitoring of dredged disposal sites has provided field confirmation of their ability to meet narrative and biological goals for sediments.

1.2.2 NO-EFFECTS CONCENTRATIONS

Overall Approach. No-Effects Concentrations (NECs) were calculated in 1996 by EPA Region 5's Great Lakes National Program Office (GLNPO) in cooperation with the National Biological Service. This approach is very similar to the calculation of AETs described above, except that it employs some additional data screening steps (see Appendix B), and compares the test sample to a control sediment instead of a reference sediment for purposes of defining an adverse effect. At the time these SQVs were derived, no freshwater AETs had yet been published.

Geographic Scope. GLNPO used a data set comprised of 83 samples from three Great Lakes areas of concern and from several rivers and estuaries, including Waukegan Harbor, IL; upper Mississippi River near Minneapolis, MN; upper Clark Fork River, MT; Trinity River near Dallas, TX; Mobile Bay, AL; and Galveston Bay, TX. No data from the Pacific Northwest were included.

Normalization and Units. All NECs are reported in dry weight units. Seven metals (As, Cd, Cr, Cu, Pb, Ni, and Zn) are reported in mg/kg, and 13 individual PAHs, LPAH, HPAH, total PAHs, and total PCBs are reported in µg/kg. AVS is reported in µm/g and TOC in percent. NECs were not calculated for other chemicals.

Biological Tests Included. NECs were calculated separately for three sediment bioassays: 1) 14-day *Chironomus riparius* survival and growth, 2) 14-day *Hyalella azteca* survival, growth, and maturation and 3) 28-day *Hyalella azteca* survival, growth, and maturation. No benthic community data were included.

Regulatory Use. These SQVs were developed by GLNPO for informational purposes, intended as guidance for use in evaluating contaminated sediment. The report briefly suggests that SQVs in general can be used to:

- Interpret historic chemistry data
- Identify chemicals or areas of concern
- Identify the need for more detailed studies before action is taken
- Identify a potential problem before discharging a chemical
- Establish a link between a source and sediment contamination
- Trigger regulatory action
- Establish target remediation objectives

However, this report was intended mainly as a demonstration of how to calculate and evaluate SQVs, and does not discuss any of these potential uses in detail. These SQVs are not currently endorsed or adopted as criteria by any agency, state, or province in North America.

1.2.3 EFFECTS RANGE LOW AND EFFECTS RANGE MEDIAN

Overall Approach. Effects Range Low (ERL) and Effects Range Median (ERM) values were originally developed as informal criteria for marine and estuarine sediments by the NOAA Status and Trends Program, using NOAA's Biological Effects Database for Sediments (BEDS). As part of the same report described above for NECs, GLNPO used a similar method to calculate ERLs and ERMs for freshwater sediments.

Using this approach, the database is screened to identify data that appear to have a concordance between elevated chemical concentrations and biological effects, and these effects data are arranged in increasing concentration. The ERL corresponds to the 15th percentile of the effects distribution, and is conceptually defined by NOAA as the concentration below which adverse effects are rarely observed among sensitive species. Although NOAA originally used the 10th percentile to calculate ERLs for marine sediments, GLNPO chose to use the 15th percentile to reduce the frequency of Type II errors. The ERM corresponds to the 50th percentile of the effects distribution, and was conceptually defined by NOAA as the concentration above which effects are frequently observed among most species. In between these two values, a range of effects may be observed. Additional details on calculation methods are provided in Appendix C.

Geographic Scope. The same data set described for NECs was used to derive ERLs/ERMs. Of these, approximately 35% were samples exhibiting adverse effects and were used in the calculations.

Normalization and Units. All ERLs and ERMs are reported in dry weight units. Seven metals (As, Cd, Cr, Cu, Pb, Ni, and Zn) are reported in mg/kg, and 13 individual PAHs, LPAH, HPAH, total PAHs, and total PCBs are reported in µg/kg. AVS is reported in µm/g and TOC in percent. ERLs/ERMs were not calculated for other chemicals.

Biological Tests Included. ERLs/ERMs were calculated separately for three sediment bioassays: 1) 14-day *Chironomus riparius* survival and growth, 2) 14-day *Hyalella azteca* survival, growth, and maturation and 3) 28-day *Hyalella azteca* survival, growth, and maturation. No benthic community data were included.

Regulatory Use. These SQVs were developed by GLNPO for informational purposes, intended as guidance for use in evaluating contaminated sediment (see above). However, they are not currently endorsed or adopted as criteria by any agency, state, or province in North America. Use of the ERLs and ERMs has generally declined over time in favor of the TELs/PELs, described below.

1.2.4 THRESHOLD EFFECTS LEVEL AND PROBABLE EFFECTS LEVEL

Overall Approach. Threshold Effects Levels (TELs) and Probable Effects Levels (PELs) for freshwater sediments were calculated by GLNPO as part of the same effort described above for NECs and ERLs/ERMs. In addition, Environment Canada has calculated TELs/PELs for marine and freshwater sediments, which have been adopted as national criteria by the Canadian Council of Ministers of the Environment.

TELs/PELs are calculated similarly to ERLs and ERMs, except that they make use of both the effects and no-effects data distributions. First, the data are screened (see Appendix D) and assigned to effects or no-effects distributions, which are each arranged in ascending order of concentration. The TELs were derived by calculating the geometric mean of the 15th percentile of the effects data set and the 50th

percentile of the no-effects data set, and is narratively described as the level below which adverse biological effects rarely occur. The PELs were derived by calculating the geometric mean of the 50th percentile of the effects data set and the 85th percentile of the no-effects data set, and is narratively described as the level above which adverse biological effects frequently occur. Between these two levels, varying levels of adverse effects can occur.

Geographic Scope. Two versions of TELs/PELs have been calculated for freshwater sediments, both in 1996. USGS calculated TELs/PELs for GLNPO using a more limited data set described in Section 1.2.3 above, focusing on Great Lakes Areas of Concern. At the same time, Environment Canada and its contractors calculated TELs/PELs from the much larger NOAA BEDS database, which included data from all over North America, including the GLNPO data, the data used by Ontario to calculate SLCs, and many other studies, including a few studies from the Columbia Basin in eastern Washington. This second set of TELs/PELs is the one in most widespread use and will be focused on here.

Normalization and Units. TELs/PELs for 8 metals (As, Cd, Cr, Cu, Pb, Hg, Ni, Zn) are reported in mg/kg dry weight, and TELs/PELs for 12 individual PAHs, total PCBs, and 7 pesticides are reported in µg/kg dry weight. For their more limited data set, USGS calculated TELs/PELs in organic carbon-normalized units and found that the dry weight TELs/PELs were equally or more reliable compared to the organic carbon-normalized values, and therefore the decision was made to use the dry weight values. Also, not all the data in the national BEDS database had TOC values with which to conduct normalization.

Biological Tests Included. The BEDS database includes toxicity values associated with a wide variety of biological tests and SQVs, including acute and chronic bioassays, benthic community studies, spiked-sediment bioassays, equilibrium partitioning values, and SQVs from other jurisdictions, such as the Ontario SLCs. More than 90% of the studies are field studies. Individual biological tests or species represented in the database vary by chemical, but generally include benthic richness and abundance (total or taxa); *Hyalella azteca* 96-hr, 10-day, 14-day, and 28-day bioassays; *Diporeia* sp. 3-day, 6-day, 12-day, 19-day, and 26-day bioassays, *Daphnia pulex* 96-hr and *D. magna* 48-hr, 6-day, 16-day, and 22-day bioassays; *Chironomus tentans* 48-hr and 10-day bioassays and *C. riparius* 10-day and 14-day bioassays; *Hexagenia limbata* 96-hr and 10-day bioassays; *Lumbriculus variagata* 30-day bioassay; and a few bioassays with fish, crayfish, snails or frogs of various species. All of the values associated with these biological test are combined in calculating the TELs/PELs.

Regulatory Use. Freshwater and marine TELs/PELs calculated as described above have been adopted as Interim Sediment Quality Guidelines by Environment Canada, and are used as freshwater criteria in Canadian provinces that do not have their own criteria. Adoption as a federal Sediment Quality Guideline in Canada means that these values are recommended to the provinces as nationally consistent benchmark values, used for screening purposes. They may be used as site-specific cleanup objectives, or the site-specific objectives may vary from these values. Between the TEL and PEL (and sometimes above the PEL), biological assessment tools are considered very important in establishing what action, if any, is needed in a particular case. The focus of their use is generally on cleanup sites; however, the need for source control could arise as one of the management options for a site.

British Columbia is in the process of updating these values for use as marine and freshwater sediment criteria by the Ministry of Water, Lands, and Parks, but the updated values are not yet available. Rather than the TEL, which is considered too conservative for cleanup sites, British Columbia currently proposes to use the Average Effects Level (AEL) as a lower level, which is the average of the TEL and PEL. Once finalized, the criteria will be used by British Columbia only as part of the Contaminated Sites program,

not for other purposes. The criteria may be used in all stages of the cleanup process, from site identification to use as legally binding cleanup standards.

Florida uses marine TELs/PELs as sediment quality assessment guidelines. These guidelines are used to help focus monitoring programs, identify areas and chemicals of concern, and indicate where additional biological studies may be needed to confirm toxicity. In addition, the freshwater and marine TELs/PELs have been included in risk assessment guidance documents in various states and provinces as benchmark values.

1.2.5 CONSENSUS-BASED SEDIMENT QUALITY GUIDELINES

Overall Approach. Consensus-based sediment quality guidelines were developed by a group of private and agency sediment researchers and colleagues in an attempt to unify the wide variety of SQVs available in the literature. All existing field-derived freshwater SQVs were collected from around North America and divided into two groups – a lower group representing levels below which effects would not be expected, and an upper group representing levels above with adverse effects would be expected. The geometric mean of the lower group was calculated and is referred to as the Threshold Effects Concentration (TEC). Likewise, the geometric mean of the upper group was calculated and is referred to as the Probable Effects Concentration (PEC). A list of these guidelines is provided in Appendix E.

The likelihood of toxicity is determined by dividing chemical concentrations by the TEC or PEC to obtain a quotient for each chemical, and then determining the mean TEC or PEC quotient. The mean PEC quotient is highly correlated to the observed incidence of toxicity in several widely-used freshwater bioassays, including acute and chronic *Hyalella azteca* and acute *Chironomus* sp. bioassays. One key aspect of this method is that the authors have demonstrated that summed PAH and PCB values provide as accurate a prediction of toxicity as do individual PAH and PCB values.

Geographic Scope. The consensus-based guidelines were not developed using an independent data set, but rather from pre-existing SQV sets. These SQV sets were developed using data from all over North America, including some from Washington State used to develop AETs and TELs/PELs; descriptions of the underlying data sets can be found in the sections for each individual SQV set.

Normalization and Units. Consensus-based guidelines are reported in dry weight, 8 metals (As, Cd, Cr, Cu, Pb, Hg, Ni, Zn) in mg/kg and organics in µg/kg. Organic chemicals include 10 individual PAHs, total PAHs, 9 pesticides, and total PCBs. Those SQVs that were originally reported in organic carbon-normalized values were converted to dry weight using an assumption of 1% TOC.

Biological Tests Included. As discussed above, the consensus-based SQVs were not developed using an independent data set. They include all the biological tests contained within the underlying data sets, which are described in the sections for each individual SQV set.

Regulatory Use. Consensus-based guidelines have been collaboratively developed and widely promoted by a group of agency and consulting scientists in North America, but are relatively recent in derivation. To date, consensus-based SQVs have been developed and included in a guidance manual as sediment quality targets for the St. Louis Area of Concern by GLNPO and the Minnesota Pollution Control Agency. The recommended applications of these SQVs to the St. Louis Area of Concern include:

- Designing monitoring programs
- Interpreting sediment chemistry data
- Assessing risks to benthic community organisms

- Developing site-specific remediation targets (in conjunction with biological effects data)

1.2.6 SCREENING LEVEL CONCENTRATIONS

Overall Approach. The Screening Level Concentration (SLC) approach was developed by the Ontario Ministry of the Environment, and is based on the presence and absence of benthic species in freshwater sediments (see Appendix F). First, a field database is gathered with synoptic chemical and benthic community data. A chemical concentration distribution is prepared for each benthic species and each chemical, using only the stations at which that species was observed. For each distribution, the 90th percentile is determined. This concentration is assumed to represent a conservative estimate of the upper tolerance level for that species and that chemical, since above that level the species is seldom observed. For each chemical, the tolerance levels of all the species are plotted on a graph by increasing concentration. From this distribution, various levels can be selected, depending on what percent of the species you wish to protect. The most widely used values, developed by Ontario for use in the Great Lakes, include the Lowest Effect Level (5th percentile) and the Severe Effect Level (95th percentile). The LEL corresponds to a level at which you would expect to see effects in only 5% of benthic species, while the SEL represents a level at which you would expect to see effects in 95% of benthic species.

Geographic Scope. The data set behind the Ontario values consists largely of samples from Lake Ontario and Lake Huron, and various rivers in Ontario that are tributaries to the lakes, along with some data from river tributaries in Michigan and New York. The number of data points varies by chemical, and is not stated in any of the references that could be readily obtained. However, there were at least 200 data points for metals, and at least 100 for organic chemicals.

Normalization and Units. In the original report, LELs are all in mg/kg, although in this report LELs for organic chemicals have been converted to µg/kg for ease of comparison to other SQV sets. SELs were originally reported as a value which, when multiplied by the TOC associated with the sample, gives the guideline value for that sample. Because this resulting level varies according to the TOC in an individual sample and therefore cannot be used in a reliability analysis, these levels have been converted to dry weight values for this report by assuming a TOC level of 1%. LELs/SELs are available for 8 metals (As, Cd, Cr, Cu, Pb, Hg, Ni, Zn), nitrogen, phosphorus, TOC, 12 individual PAHs, total PAHs, 14 pesticides, 4 individual Aroclor mixtures, and total PCBs.

Biological Tests Included. All samples included benthic community data, down to species level (where possible). Only the presence or absence of individual species was considered. 100 species were selected for the SLC calculations that were considered representative of the entire benthic community, including both pollution-tolerant and pollution-sensitive species.

Regulatory Use. Ontario uses these SQVs as guidelines for their dredging, cleanup, and source control programs. Below the LEL, sediments are not considered to require action and may be disposed of in open water. Between the LEL and the SEL, biological testing is used to determine the level of toxicity and develop an appropriate management plan for the sediments. Sediments at this level of contamination may only be disposed of in areas that are similarly degraded. Above the SEL, the likelihood that cleanup will be required increases, and open-water disposal is prohibited.

Quebec also used this approach to establish SQVs for management of dredged material and cleanup of contaminated sites in the St. Lawrence River. However, they used a database with values for the St. Lawrence River and modified the percentiles to the 15th percentile, known as the Minimum Effects Threshold (MET), and the 90th percentile, known as the Toxic Effect Threshold (TET). The Ontario values were developed earlier and are much more widely used as freshwater SQVs by other states and

provinces in the Great Lakes region, and therefore these values are used in this report in preference to the Quebec values.

1.2.7 EQUILIBRIUM PARTITIONING

Overall Approach. The Equilibrium Partitioning (EqP) method, unlike the other methods described above, is based on partitioning theory and laboratory measurements of toxicity of chemicals in water to various freshwater aquatic species. This method has been sponsored and selected by EPA headquarters as its primary approach to national SQVs. One of its primary assumptions is that the toxicity of a chemical in sediments is equal to its toxicity in water, multiplied by a sediment/water partitioning coefficient. Therefore, the SQVs developed under this approach are based on the Final Chronic Value for water, which is the water quality criterion developed to be protective of 95% of freshwater species in chronic laboratory toxicity tests. This Final Chronic Value is multiplied by the sediment/water partitioning coefficient to derive the SQV.

The most difficult step in this model is determining the partitioning coefficient, and this is where most of EPA's development work has taken place. For nonionic organic chemicals, the partitioning coefficient is assumed to be equal to the K_{oc}, which can be estimated from laboratory measurements of K_{ow} through a linear equation relating the two. This approach has been used to develop EqP values for 34 nonionic compounds, including pesticides and herbicides, chlorinated hydrocarbons, phthalates, BTEX compounds, and other miscellaneous organic chemicals. In addition, an SQV has been developed for PAH mixtures, based on principles of narcosis toxicology, in which the toxicity of PAHs and other nonionic organics has been shown to be additive on a molar concentration basis.

For metals and ionic organic compounds, the relationship is more complex and site-specific. The partitioning coefficient for metals and ionic organic compounds is usually referred to as K_D, which is a bulk sediment/water partitioning coefficient. This partitioning coefficient is difficult to predict from intrinsic chemical properties, and generally must be measured in the field. EPA has not yet developed EqP values for any ionic organic chemicals.

For metals, EPA has developed an SQV for mixtures of six metals – cadmium, copper, lead, nickel, silver, and zinc based on the theory that these metals can be bound by acid-volatile sulfides (AVS) in sediments, which renders them unavailable for partitioning. For these metals, the total concentration of the simultaneously-extracted metals (SEM) is subtracted from the concentration of AVS in sediments. If AVS minus SEM is greater than zero, the metals are believed to be not bioavailable and non-toxic.

Geographic Scope. The EqP method is derived entirely from partitioning theory and toxicological models, which are in turn based on laboratory measurements of partitioning behavior and toxicity for individual chemicals. Therefore, there is no underlying field data set or any associated geographic scope. Theoretically, it is considered applicable to all freshwater environments in North America. However, the toxicity values used in the models were derived using certain species (see below), which may or may not be representative of freshwater species on a regional basis. Therefore, EPA has provided procedures for modifying the toxicity values and deriving site-specific SQVs using regional species.

Normalization and Units. SQVs for non-ionic organics are listed in mg/kg OC-normalized, including 12 pesticides/herbicides, BTEX, and 16 miscellaneous organics such as phthalates and chlorinated alkanes. The SQV for PAH mixtures is stated as a toxic unit of 1. The SQV for SEM-AVS is 0 um/g.

Biological Tests Included. Final Chronic Values are based on laboratory water toxicity tests with a variety of vertebrate and invertebrate species, including amphipods, barnacles, bivalves, snails,

cladocerans, flies, midges, various other benthic invertebrates, sea urchins, crustaceans, salmonids, other fish, and amphibians. Exact numbers and types of species vary from chemical to chemical, and are documented in the references provided in Appendix G. Both freshwater and saltwater species were used to calculate most SQVs, as partitioning of nonionic chemicals is not expected to vary according to ionic strength. However, both freshwater and saltwater values are provided for Endrin and Dieldrin. This is most likely because EqP values for these two chemicals were derived much earlier than for the other chemicals, and were not updated using the most recent approach.

Regulatory Use. A variety of EqP SQVs and guidance documents were due to be finalized by EPA in late December 2000, but publication of these documents in the Federal Register and their adoption as final criteria was indefinitely put on hold by the Bush administration. The current EqP values are available to the public, but are considered draft and not official EPA guidance. EPA's implementation guidance document suggests that these SQVs could be used in various programs to:

- Develop water quality criteria for the protection of sediment quality
- List water bodies as water quality-limited due to sediment toxicity
- Develop TMDLs for such water bodies
- Establish NPDES permit conditions for discharges
- Determine the suitability of dredged material for open-water disposal
- For use in establishing sediment cleanup objectives under Superfund and RCRA

The New York Department of Environmental Conservation is the only state or province to have used the EqP approach as the basis for its sediment guidelines, which it did in 1993. These values are primarily considered screening concentrations for the site cleanup program, although they could be used as remediation targets for smaller sites. Although NYDEC developed values for many more chemicals than EPA had available at that time, the current EPA draft values are more up-to-date and reflect the best available science, which has progressed considerably since 1993. Therefore, the EPA values are used in this document in preference to the NYDEC values.

1.3 SQV Set Comparisons and Recommendations

In this section, the SQV sets that were compiled and described in Section 1.2 were evaluated against a set of criteria to determine which ones should be carried forward for reliability analysis. These criteria are also used to identify those existing SQV sets that appear to be the most technically sound and compatible with the narrative goals of Ecology's regulatory programs. This information will later be combined with the reliability assessment conducted in Task 5 to provide recommendations to Ecology on the use of these SQV sets in Ecology's freshwater sediment programs. Ecology will prepare final guidance on how these values should be used.

The following are the criteria that were used to evaluate the SQV sets. The first six criteria were used to select SQVs sets for the reliability analysis. Criterion 6 is included in this assessment only because it is important to compare potential Washington State methods with those that are in widespread use in North America, even if some of those SQV sets do not score highly under some of the other criteria. Criteria 1-5 and 7 will be used under Task 6 to prepare recommendations for using the SQV sets as part of Ecology's programs.

1. **Consistency with Ecology's Sediment Management Standards.** Were the SQVs designed to protect against both acute and chronic effects and do they actually do so in practice? Is their narrative or policy goal consistent with "no adverse effects" and/or "minor adverse effects"? Is the method similar to SQV methods in use for marine sediments in Washington State?

2. **Technical merits.** Is the approach internally consistent and mathematically defensible? Is the data set of high quality and what degree of quality assurance was conducted? Has the approach been reliability-tested and/or field-verified? Has the approach been published and/or validated by independent scientific review?

3. **Applicability to field conditions.** Does the method incorporate the influence of chemical mixtures in sediments? Does the method incorporate a wide range of biological effects and ecosystem niches? Does the method include direct measures of biological effects, including *in situ* effects?

4. **Biological relevance.** Are the biological tests used to develop the criteria relevant to organisms or benthic communities indigenous to the Pacific Northwest?

5. **Practicability.** Does the method currently have SQVs for a wide range of chemicals that would be expected in Washington State freshwater sediments?

6. **Regulatory Use and Representativeness.** Is the SQV development approach in widespread use in North America? Is this SQV approach unique or the best representative of a group of similar SQV sets, or can it be represented by another set of SQVs that scores more highly against the above criteria?

7. **Reliability in predicting adverse effects in freshwater sediments in Washington State.** This criterion will be assessed once Task 5 is completed, and includes assessment of false negative rates, false positive rates, and overall reliability of the SQV sets in predicting adverse biological effects in existing data sets from Washington State.

Evaluation for Reliability Analysis

Table 1-1 summarizes the scores that were assigned to each of the SQV sets using Criteria 1-6 described above. The purpose of this assessment was to select a subset of the available SQV sets to be carried forward for reliability assessment, not to select a single set for use in Washington State. Therefore, the individual scores are not quantitative, nor was it considered important to rigorously assess each variable. Scores were developed by a group of agency staff and contractors with expertise in SQVs, generally by consensus. There were some areas where scores would have varied slightly from person to person. However, even taking these differences into account, the relative scores remained the same among the SQV sets, and the conclusions regarding which ones to retain for reliability analysis did not change.

Table 1-1. Summary of SQV Scores for Criteria 1-6

Criterion	AET	PAET	NEC	ERL	ERM	TEL	PEL	TEC	PEC	LEL	SEL	EqP
1. Consistency w/SMS	+	+	++	+	-	+	-	+	-	++	-	+
2. Technical Merits	0	0	+	+	+	0	0	+	+	-	-	-
3. Field Applicability	+	+	+	+	+	++	++	+	+	++	++	--
4. Biological Relevance	+	+	+	+	+	+	+	+	+	-	-	--
5. Practicability	++	++	0	0	0	+	+	+	+	++	++	--
6. Regulatory Use	+	0	-	0	0	++	++	+	+	++	++	+
TOTAL SCORE:	6	5	4	4	2	7	5	6	4	6	3	-5

Scores	Definitions (see text for further explanation)
++	Highest performing SQVs; meets all criteria
+	Above average; meets most criteria
0	Average; meets some criteria but not others
-	Below average; meets few criteria
--	Lowest performing SQVs; meets none of the criteria
	SQV sets that will be retained for reliability analysis

1.3.1 CONSISTENCY WITH ECOLOGY'S SEDIMENT MANAGEMENT STANDARDS

SQV sets received one plus in this category if their narrative policy goal was between "no adverse effects" and "minor adverse effects", if the derivation method is consistent with the AETs already in use under the Sediment Management Standards, ***or*** if the SQVs currently available incorporate both acute and chronic effects. If all of these criteria are met, the SQV set received two pluses. If only one or none of these criteria are met, the SQV set received a minus score.

AETs received one plus because there are currently AETs and PAETs for only two bioassays, and very little chronic data were available at the time they were calculated. NECs, which are very similar to AETs, received two pluses because they incorporated a significant amount of chronic data and NECs were calculated for three bioassays. ERLs, TECs, and PECs all received one plus because they include some chronic data and have narrative policy goals similar to the SMS, although their methods of derivation are quite different. The higher SQV in each of these sets, ERMs, PELs, and PECs, all received minus scores because the narrative policy goal for these criteria allows more than minor adverse effects. LELs received two pluses, because on a species basis the method of derivation is very similar to AETs, chronic (benthic) data are included, and the narrative policy goal falls within the SMS range. Like the other SQV sets, however, the higher SQV in this pair, the SEL, received a minus score because it corresponds to a severe effects level. Finally, EqP received a plus score because it incorporates chronic effects and has a narrative policy goal within the SMS range, but has a very different method of derivation.

1.3.2 TECHNICAL MERITS

Scoring under this category was largely based on the degree of quality assurance that was conducted and the extent of reliability testing, peer review, and publication the method has received. AETs have received extensive peer review and reliability testing, but the development of the preliminary freshwater guidelines did not incorporate rigorous quality assurance. Also, there was sufficient data to calculate AETs for only two species, fewer than would be considered sufficient to represent a cross-section of the benthic community. Perhaps more importantly, no benthic community data and very little chronic data was available. TELs/PELs had a more careful quality review of the initial data, but suffered from other problems, including incorporation of a wide variety of different types of values (different biological tests and derivation methods) into one database, inconsistent hit/no-hit definitions, and adoption of some values with low reliability. These SQVs received a score of zero.

NECs and ERL/ERM values calculated by USGS/GLNPO received a relatively high level of quality assurance and were subjected to a sophisticated reliability assessment. These values were peer-reviewed and published in both agency and peer-reviewed publications, and distributed widely in North America. The suite of bioassays used in the analysis included two chronic bioassays, as well as acute bioassays. The same group of agency researchers and consultants has more recently been developing and conducting reliability assessments on the TEC/PEC values, with impressive reliability results. These SQV sets received plus scores.

The LEL/SEL values received minus scores largely because quality assurance procedures could not be verified, and the values were calculated before many current protocols existed. No reliability assessments of these values have been conducted, and it is unknown whether confounding factors such as physical effects may have influenced the presence or absence of species. Finally, EqP values also received a negative score. This approach has been subjected to significant peer and scientific review, yet a scientific consensus has yet to emerge on its usefulness and applicability to field conditions. EqP values have not yet been field-verified or reliability tested against field data.

1.3.3 APPLICABILITY TO FIELD CONDITIONS

SQV sets received a plus score in this category if they were derived entirely or largely from field-collected, synoptic chemistry and bioassay data, incorporating the effects of mixtures and field conditions on toxicity. Most SQV sets are in this category, including the AETs/PAETs, NECs, ERLs/ERMs, and TECs/PECs (which themselves are based on the other SQVs listed). A double-plus score was assigned to SQV sets that made significant use of benthic data, primarily the Ontario LELs/SELs. The TELs/PELs also received a double-plus because they incorporated the Ontario benthic data. EqP values received a double-minus because no field data were used to derive these values (except in the case of the PAH narcosis model), and there has been almost no field-verification of the models.

1.3.4 BIOLOGICAL RELEVANCE.

There is essentially no difference among most of the SQV sets with respect to biological relevance to the Pacific Northwest, because all of these SQV sets were derived using a similar and limited set of bioassays. For this reason, AETs/PAETs, NECs, ERLs/ERMs, PELs/TELs, and TECs/PECs all received a plus score. The Ontario LELs/SELs received a minus score because this method is based entirely on the presence or absence of Great Lakes benthic species, and it is anticipated that freshwater benthic assemblages in Washington State would likely be significantly different from those present in the Great Lakes. Similarly, the EqP values are based primarily on toxicity to water column organisms, and the reliability of these values in predicting effects to benthic species in the field has not been demonstrated. Therefore, EqP received a double-negative score.

1.3.5 PRACTICABILITY

AETs/PAETs and LELs/SELs have SQVs for the greatest number of sediment chemicals of concern in Washington State, and would be the most practical to use at this time for that reason. These SQV sets received a double-plus score. TELs/PELs and TECs/PECs had a similar and somewhat more limited list of chemicals, and received a plus score. NECs, ERLs, and ERMs were limited to metals and PAHs, and received a zero score. EqP received a double-minus score because there are almost no SQVs available for chemicals of concern, in the form we have been measuring them. Most of the data collected in Washington State does not include AVS/SEM measurements, and even if this were available, there are several other metals of concern. The narcosis-based PAH approach requires analysis of over 30 PAH compounds, and very few data sets have the required data to calculate the sum PAH value. Although there are EqP values for quite a variety of other miscellaneous organic compounds, most of these have not been detected in Washington State sediments and are not considered chemicals of concern.

1.3.6 REGULATORY USE AND REPRESENTATIVENESS

SQV sets received a double-plus score in this category if they are among the most widely-used SQVs for freshwater sediments in North America, and TELs/PELs as well as the Ontario LELs/SELs qualified for this score. SQV sets received a plus score if they have been used in at least one jurisdiction in North America. AETs, TECs/PECs, and EqP qualify for this score. SQV sets received a zero score if they are a modification of an approach used by at least one jurisdiction, and PAETs, and ERLs/ERMs (as calculated by USGS) qualified for this score. NECs have not been used by any jurisdiction and received a minus score.

1.4 Overall Comparison of SQV Sets

In comparing the final scores for the SQV sets, it is useful to compare scores for similar approaches in addition to the overall scores. NECs are very similar in derivation to AETs/PAETs, but received a lower

score. Therefore, NECs are screened out and AETs/PAETs are retained. Similarly, ERLs/ERMs are similar to PELs/TELs, yet PELs/TELs received a higher score. Therefore, ERLs/ERMs are screened out and PELs/TELs are retained. TECs/PECs and LELs/SELs both received sufficiently high scores to be retained as well, and represent different approaches from the other two sets. EqP is screened out because it received a very low score compared to the other SQV sets. In addition, it would be very difficult to conduct a reliability assessment of EqP values because the information needed to compare SEDQUAL data against the EqP values is largely missing, as discussed above.

1.5 Approaches to Deriving SQVs Using Benthic Community Data

Inclusion of benthic community data has been an important element of Washington State's approach to developing SQVs since the AETs were first calculated in 1986. As a component of the marine standards, benthic community AETs provide a means of ground-truthing laboratory bioassay results and ensure that chronic effects are measured and included in Ecology's regulatory framework. As part of the "sediment quality triad", chemistry, bioassays, and benthic community data each have their role to play in creating a scientifically defensible association between chemical elevations in sediments and population-level effects in the field. Both the Department of Ecology and EPA Region 10 have expressed a strong commitment to including benthic community data in the derivation of freshwater sediment quality guidelines, if at all feasible. The inclusion of benthic community data would be especially helpful in the freshwater arena, where chronic freshwater bioassays have only recently become available and few chronic data currently exist. Even acute bioassays are relatively few in number, leading to concerns that sensitive species within the benthic community may not be adequately represented or protected.

The Ontaro LELs/SELs, described above, are derived using benthic data and could potentially be used. These SQV sets have been carried forward for reliability analysis. However, they suffer from a variety of issues that may make them unsuitable for use in Washington State. No quality assurance information is available for them, there has been no field-testing to evaluate the reliability of these criteria, and they are based on Great Lakes species assemblages, which may not be relevant to freshwater ecosystems in Washington State. Nevertheless, they could serve as a benchmark comparison to evaluate the general sensitivity of SQVs derived using bioassay data.

Described below are several methods that could be used to derive Washington State SQVs using benthic community data. Use of any of these methods is dependent on having a large and diverse regional benthic data set, coupled with synoptic chemistry data, including a range of clean to contaminated areas. Although a substantial benthic database exists, very little chemistry data has been collected along with it. Currently, there are approximately only 10 surveys that include both benthic and chemistry data. Unlike marine ecosystems, it is also possible that freshwater ecosystems in Washington State will prove to be more diverse and will require stratification of the data into ecoregions, increasing overall data requirements.

1.5.1 Apparent Effects Thresholds

Similar to the benthic AET that currently exists for marine sediments, a freshwater AET could be developed for benthic community impacts. This AET would be included along with other acute and chronic AETs based on bioassay endpoints to determine the lowest and second-lowest freshwater AET. The advantage of this approach is that it would fit into the regulatory program already in place for marine sediments and would add the benthic community leg of the triad that was envisioned in the development of AETs.

As has been the case for the marine AET, the greatest difficulty in pursuing this approach may be selecting benthic endpoints and defining an adverse impact. Many different endpoints exist, and there has been considerable discussion for the last 10 years about which ones are the most appropriate to use. The benthic endpoints originally defined for the marine AETs are currently being revised based on recent recommendations from regional experts. While this work may also have some utility in selecting freshwater endpoints, certain endpoints that are taxa-specific would not be relevant to the freshwater environment (e.g., polychaete abundance). However, analogous endpoints (e.g., oligochaete abundance) could possibly be developed.

1.5.2 ONTARIO SLC APPROACH

The Ontario SLC approach is the only method that has been used so far in the US or Canada to develop freshwater SQVs with benthic data. This method could be used with a Pacific NW regional benthic data set to develop benthic SQVs appropriate to Washington State. The SLC approach is conceptually similar to the AETs in that an upper tolerance threshold is derived for each species to a chemical. These tolerance thresholds are arranged in increasing order of concentration and a low-end value is selected that is protective of most species. The primary difference is that there are many more benthic species than there are biological tests used to develop AETs.

One advantage of this approach is that it uses presence or absence of each species to define effects thresholds rather than more complex (and controversial) measures of benthic community effects. In addition, using benthic community data provides a more direct link to field effects than do the bioassay endpoints typically used to develop most AETs. However, one drawback to the SLC approach is that confounding factors such as fines that may affect the presence or absence of a species are not taken into account. Covariance of fines with species presence or absence, as well as with chemical concentrations, should be evaluated if this method is selected for use.

1.5.3 BENTHIC INDEX OF BIOLOGICAL INTEGRITY

The Benthic Index of Biological Integrity (B-IBI) is used by the Washington Department of Ecology and several other agencies to monitor the health of freshwater ecosystems under the Clean Water Act. The index adds together scores for a number of different measures to arrive at an overall score for each station ranging from 9 to 45, as shown in Table 1-2. These criteria are currently calibrated for the ecoregion of the state known as the Puget Lowlands and would need to be recalibrated for other ecoregions.

The Washington Department of Ecology uses scoring ranges shown in Table 1-3 to describe benthic community health. These ranges could be used as an endpoint for developing AETs or other SQVs. Another approach might be to develop regressions of the IBI against chemical concentrations to identify SQVs. This latter approach would be complicated by chemical mixtures and the effects of physical parameters, however.

Table 1-2. Metrics included in the Benthic Index of Biological Integrity (B-IBI)

		Scoring Criteria		
Metric	**Response**	**1**	**3**	**5**
Total number of taxa	Decrease	< 10	10 - 20	> 20
Number of Ephemeroptera taxa	Decrease	< 3	3 - 5.5	> 5.5
Number of Plecoptera taxa	Decrease	< 3	3 - 5.5	> 5.5
Number of Trichoptera taxa	Decrease	< 2	2 - 4.5	> 4.5
Number of long-lived taxa	Decrease	< 0.5	0.2 - 2	> 2
Number of intolerant taxa	Decrease	< 0.5	0.5 - 2	> 2
% of individuals in tolerant taxa	Increase	> 50	20 - 50	< 20
% of predator individuals	Decrease	< 5	5 - 10	> 10
% dominance (2-3 taxa)	Increase	> 75	50 - 75	< 50

Table 1-3. Scoring Ranges for the B-IBI Index

Score	Classification	Description
33 – 45	Good	Natural biological conditions
21 – 33	Fair	Slight impairment of biological conditions
9 – 21	Poor	Obvious impairment of biological conditions

1.5.4 MULTIVARIATE REFERENCE RANGE APPROACH

The reference range approach was developed by the National Water Research Institute for the International Joint Commission on the Great Lakes, for use in assessing sediment quality at the Great Lakes Areas of Concern (Reynoldson et al., 1997; Reynoldson and Day, undated). This is a complex approach based on a variety of chemical and biological measures that can be numerically evaluated and input into a multivariate model.

Under this approach, reference areas within the Great Lakes have been selected and exhaustively characterized, using a variety of physical and chemical parameters, toxicity tests, and benthic community analyses. From these studies, a numeric "reference condition" has been established for each parameter and entered into a database/analytical software package called BEAST – Benthic Environmental Assessment Tool. A mathematical model has been developed relating these parameters to environmental impairment in the Great Lakes, which will be used to assess sediment conditions at areas of concern. Environmental data measured at the AOCs will be entered into the model, and using a multivariate analysis, it will determine how different from the reference condition the site is. Standard deviations are used as the measure of difference, where 2-3 standard deviations from the reference mean would be considered heavily impacted and likely in need of cleanup or restoration. This approach is viewed as an alternative to setting "bright line" SQVs, and is favored by some in the US as well as Canada. However, it is research- and data-intensive to conduct the initial reference area evaluations and develop the model, and is highly region-specific.

In order to use this approach in Washington, we would need to establish freshwater reference areas, which could be significantly more difficult than in the Great Lakes, where conditions are relatively more homogeneous. A database would need to be developed of chemical, physical, and biological conditions in the reference areas, and a model developed to relate the chemical and physical parameters to benthic

community health. Because this tool was developed as an alternative to chemical SQVs, it is not immediately apparent how one would use it to develop such values. However, it is possible that boundaries representing standard deviations from the mean reference condition could be used to identify chemical values associated with those boundaries.

1.5.5 RIVPACS

Under a US Forest Service and US EPA grant, Dr. Charles Hawkins of Utah State University is developing a Pacific NW version of the River Invertebrate Predictive and Classification System (RIVPACS) for use in freshwater biological monitoring and watershed assessment. This approach is similar to BEAST, as it is a multivariate model calibrated to regional data, but has been developed for streams and small rivers (it may not be appropriate for lakes and large rivers or could require additional development). The model was first developed in the UK, and has also been used in Australia and the Rocky Mountain area.

This approach uses a variety of habitat and other variables to model stream systems and predict the benthic community assemblage that would be expected in a reference stream in the absence of contamination or habitat alteration. 225 reference streams have already been monitored in eastern and western Washington, as well as in Oregon and other Pacific NW states, to calibrate the model. The process of calibration identifies ecosystems characterized by a combination of variables; those which appear to be important in the Pacific NW model include latitude, geographic features such as basins and ranges, and habitat features such as slope and width of the stream.

Once reference community assemblages have been determined for each type of system, streams can be assessed to determine their level of impairment. The observed taxa are compared to the expected taxa, and the result is expressed as a ratio of observed to impaired, also expressed as a percentage. A percentage score of 80% is considered slightly impaired. To date this approach has mainly been used to characterize and assess trends in stream quality. Not as much work has been done on differentiating between possible causes of impairment, such as habitat alterations vs. chemical contamination.

1.5.6 REFERENCE RANGES

Avocet Consulting and Striplin Environmental Associates used a simpler variation of the reference range approach and the Ontario approach to develop adverse effects levels and warning levels for the Capital Regional District sewage outfalls in Victoria BC, in cooperation with BC MELP (CRD 2000). This approach was selected because existing SQVs were poorly-correlated with observed adverse effects and bioassay results were difficult to interpret. Reference ranges were developed for several benthic community parameters, and chemical concentrations were graphed against the same benthic community measures. The point at which the benthic community measures dropped below two standard deviations from the reference mean was used to establish numeric warning levels for several chemicals (including TOC) that appeared to be correlated with benthic community impairment.

1.6 Summary and Recommendations

Existing SQV sets for freshwater sediments were compiled from around North America, and fell into six general categories according to their method of derivation:

- AETs and modifications, including PAETs and NECs
- ERLs and ERMs
- TELs and PELs

- Consensus-based TECs and PECs
- Ontario SLC values, and modifications used by Quebec
- EPA EqP values, and related values derived by New York

These SQV sets are summarized in the main text in Section 1.2 and described in more detail in the technical appendices. In addition, there are a variety of benthic approaches that could be used to derive SQVs, but have not yet been used for this purpose. These methods are described in Section 1.6, as there may be interest in following up on these approaches in Phase II.

The existing SQV sets were evaluated against several criteria to determine which of the available approaches should be carried forward for reliability testing under Task 5. Based on an assessment of consistency with the SMS, technical merits, field applicability, biological relevance to the Pacific Northwest, practicability of use, and regulatory use, the following SQV sets were retained for reliability testing:

- Washington State Apparent Effects Thresholds (AETs) and Probable Apparent Effects Thresholds (PAETs)
- Canadian Council of Ministers of the Environment (CCME) Threshold Effects Levels (TELs) and Probable Effects Levels (PELs)
- Consensus-based Threshold Effects Concentrations (TECs) and Probable Effects Concentrations (PECs)
- Ontario Ministry of the Environment Lowest Effects Levels (LELs) and Severe Effects Levels (SELs)

As part of Task 5 (Section 3.0 of this report), the numeric SQVs associated with each of these pairs is reliability-tested against a regional data set of synoptic chemistry, bioassay, and benthic data to determine their false negative and false positive rates, and their overall reliability in making correct predictions regarding toxicity in regional freshwater sediments. The results of this evaluation is used to make recommendations to Ecology on: 1) Which of the existing SQVs, if any, best reflects the narrative goals of the SMS and is best-suited for use in Ecology's sediment programs, and 2) whether there is a need for further development of AETs or other freshwater SQVs in Phase II.

2.0 DATA COLLECTION AND SCREENING (TASKS 3 & 4)

2.1 Introduction

This section describes the collection and screening of freshwater sediment data sets for use in the development of freshwater sediment quality values (FSQVs) for use in Washington State. Sediment data sets from Washington, Oregon, Idaho and British Columbia that contained synoptic chemistry and toxicity data were collected and entered into the Washington Department of Ecology's SEDQUAL database. These data sets, plus other data sets already contained in SEDQUAL, were screened qualitatively, and data sets considered acceptable for use in the analysis were identified. This effort encompassed Tasks 3 and 4 of Phase I of a project entitled the Development of FSQVs for Use in Washington State. The goals of these tasks included:

Task 3-Identify and Compile Available Washington State and Regional Freshwater Synoptic Data Sets

- Identify and obtain synoptic freshwater sediment chemical and bioassay data sets from Washington State, Oregon, Idaho, and British Columbia
- Develop a checklist of data types and ancillary information that will be sought from each data set

Task 4-Select an Approach and Conduct Screening of Specifies Washington State and Regional Data Sets

- Identify and apply a method for conducting a review and screening of data sets, identified in Task 3, for use in developing freshwater sediment quality values.
- Identify alternate methods for conducting data set screening, and recommend a single method to Ecology
- Update and suggest modifications to the qualitative quality assurance methods used in Creation and Analysis of Freshwater Sediment Quality Values in Washington State (Cubbage 1997).
- Apply the recommended data screening approach to determine which data sets should be included in the development of freshwater sediment quality values.

Section 2.2 of this section describes the methods used to acquire new datasets, the sources from which datasets were received, and the nature and type of data acquired. Section 2.3 presents a framework for the qualitative screening of the acquired data sets. The results of the qualitative screening evaluation is presented in Section 2.4, including a summary table of the studies reviewed and the screening evaluation results, and a summary table of the datasets that were determined to be acceptable. A list of contacts is presented in Appendix I, and Appendix J provides the completed Freshwater Sediment Screening Criteria Checklists.

2.2 Data Set Acquisition

The identification and acquisition of synoptic freshwater datasets was accomplished in a systematic manner. A standardized search procedure was established for consistency that included preparing a data-gathering checklist for each contact (Figure 2.1). The data-gathering checklist summarized general information regarding a prospective study's sponsor, purpose, design, supporting documentation, and data availability. Initial determinations for acquisition were based on whether the available dataset contained synoptic freshwater sediment chemistry and toxicity data. A secondary determination, based on the relative completeness and availability of the dataset, was made to determine whether to proceed in obtaining a given dataset and accompanying documentation for further review.

Survey Name(s) and Location(s) _____

Date of Survey(s) _____

Location of Master Copy of Survey Report/Data _____

Agency _____ Contractor Name? _____

Contact _____ Notes: _____

Phone Number _____

Date Prepared _____

Prepared By _____

SYNOPTIC DATA SET:

Are chemistry and bioassay results included? (Y/N) _____

- If NO, data set is not acceptable

Are location data (e.g., latitude/longitude) acceptable, or can they be determined from a map? (Y/N) _____

If NO, describe status of location data ("dummy" coordinates may be assigned if QA is acceptable):

_____

_____

SEDIMENT CHEMISTRY ANALYSIS METHODS:

List classes of chemical analyzed (e.g., metals, VOAs, PAHs, Pest/PCBs, etc.): _____

_____

Are standardized chemical analysis methods used (e.g., PSEP, SW-846, CLP)? (Y/N) _____

List methodology: _____

_____

If nonstandard method, is reference provided or can methods be obtained for review? (Y/N) _____

- If NO, data for that chemical class is not acceptable.

List metals extraction method: _____

- Either Strong Acid Digestion (SAD) or Total Acid Digestion (TAD) is acceptable.

SEDIMENT QA PROCEDURES:

Recommended laboratory QA/QC requirements are provided below. Indicate QA/QC conducted:

Analysis Type	Method Blanks	Replicates	CRM	Matrix Spike	Surrogates
VOAs	✍	✍		✍	✍
SVOCs	✍	✍	✍	✍	✍
Pesticides/PCBs	✍	✍	✍	✍	✍
Metals	✍	✍	✍	✍	

- Omission of some QA/QC may not necessarily result in data rejection.
Additional review may be necessary on a case by case basis.

Figure 2-1. Freshwater Sediment Data Screening Criteria

SEDIMENT QA PROCEDURES (Continued):

Indicate what QA documention reviewed (e.g., data report appendix, separate QA report):

Does QA Report (e.g., QA1) indicate significant problems? (Y/N)
If YES, list problems:

Holding times met? (Y/N)
If NO, list holding time exceedance for classes of chemicals:
-Data with gross holding time exceedances are rejected.

Acceptable detection limits? (Y/N)
-Refer to Ecology's SAPA detection limits.
If NO, list problems noted:
-Data should be appropriately qualified or data are rejected.

BIOASSAY ANALYSIS METHODS:
List bioassay tests conducted:

Modern (post-1985) or standardized bioassay protocols used (e.g., ASTM, PSEP)? (Y/N)

List methodology:

If nonstandard method, is reference provided or can methods be obtained for review? (Y/N)
- If NO, data set is not acceptable.

BIOASSAY QA PROCEDURES:
Replicate treatments used for all bioassay tests? (Y/N)
If NO, test is not acceptable - indicate what test(s):

If YES, what are the minimum number of replicates conducted per test?

Figure 2-1. Freshwater Sediment Data Screening Criteria (Continued)

BIOASSAY QA PROCEDURES (Continued):

Negative bioassay control run (minimum of one per batch)? (Y/N)
- If NO, data for that bioassay is not acceptable.
Positive bioassay control run (minimum of one per batch)? (Y/N)
- If NO, data for that bioassay is not acceptable.

Bioassay reference run? (Y/N)
- Not a criteria for acceptance or rejection.

Does QA Report (e.g., QA1) indicate significant problems? (Y/N)
If YES, list problems:

_____
_____
_____

Holding times met? (Y/N)
If NO, list holding time exceedance for bioassays:
-data with gross holding time exceedances are rejected.

_____
_____
_____

SUMMARY:

- ✍ All data are accepted.
- ✍ All data are rejected. List reasons for data rejection:

  _____
  _____
  _____

- ✍ Partial data acceptance. List rejected data classes and reason for rejection:

  _____
  _____
  _____

Priority for data entry:
- ✍ High - acceptable data
- ✍ Low - problems noted

Is there sufficient QA documentation to conduct a QA1 data review (Y/N)? _____
- Refer to data acquisition checklist.

Is there sufficient QA documentation to conduct a QA2 data review (Y/N)? _____
- Refer to data acquisition checklist.

Figure 2-1. Freshwater Sediment Data Screening Criteria (Continued)

Preliminary meetings in September 2001 with Brett Betts of Ecology and Teresa Michelsen of Avocet Consulting provided SAIC with several initial contacts. These contacts included Washington Department of Ecology (Ecology), Oregon Department of Environmental Quality (DEQ), U.S. Army Corps of Engineers Portland District (COE), EPA Region 10, U.S. Geological Survey (USGS) Tacoma, and Idaho DEQ. Some of these contacts provided datasets for review and others provided additional leads. The majority of the datasets available for review came from Ecology's EAP files in Lacey, Washington. The USGS and Oregon DEQ also provided several datasets. In addition to reviewing available data reports, electronic data from studies previously entered into the SEDQUAL database were also evaluated. The complete list of contacts made is presented in Appendix I.

Numerous contacts were made over a period of several months, however only a limited number of new datasets were identified. Many of the data sets were limited to sediment chemistry data and were therefore not relevant for this study. A total of 58 surveys were acquired, or were already available in the SEDQUAL database, and subsequently screened as described in the following sections.

2.3 Data Set Screening

The screening of freshwater datasets was a two-part process that included completing a qualitative checklist for each study reviewed, and then screening the available datasets against a set of minimal data quality requirements to determine their acceptability for inclusion in developing FSQVs. This section describes the development and contents of the checklists, the dataset screening evaluation process, and the results of the screening process.

2.3.1 Data Screening Checklists

A checklist, *Freshwater Sediment Data Screening Criteria*, was developed in cooperation with Brett Betts of Washington State Department of Ecology, and Theresa Michelson of Avocet Consulting, for evaluating data sets. The checklists were designed to provide a qualitative overview of the type of data collected, methodology used for analysis and testing, quality assurance procedures implemented, and the results of any subsequent data quality assurance/quality control (QA/QC) reviews for both sediment chemistry and toxicity data. The checklists provide a summary of the available information regarding selected data quality variables for each data set. The checklists **do not** provide a formal or rigorous assessment of data quality (i.e. QA level 1 or 2) or data validation and should not be considered a verification of results for uses other than those discussed in this document.

The screening criteria proposed for evaluation of freshwater sediment data were selected based on a review of the qualitative screening methods and quantitative quality assurance criteria provided in the following documents:

- Creation and Analysis of Freshwater Sediment Quality Values in Washington State (Cubbage et. al., 1997)
- Sediment Sampling and Analysis Plan Appendix: Guidance on the Development of Sediment Sampling and Analysis Plans Meeting the Requirements of the Sediment Management Standards—Chapter 173-204 WAC. (Ecology 1995)
- PSDDA Guidance Manual for Data Quality Evaluation for Dredged Material Disposal Projects (PTI 1989)

The checklists were completed to the fullest extent based on the documentation available for review. Due to the variable nature of the documentation, not all checklist entries could be answered unequivocally. Incomplete or missing documentation regarding specific details of a given survey was a frequent

difficulty. Every effort was made to indicate any uncertainties regarding the data types, methodology, and overall quality of the datasets reviewed. The completed checklists for all studies reviewed are presented in Appendix J.

2.4 Evaluation of Data Sets

The evaluation of the freshwater sediment datasets consisted of the qualitative screening of the available data sets. Two components of the data sets were evaluated: 1) a set of minimal data quality requirements; and 2) a more ambiguous component of additional QA considerations based on significant QA/QC problems and/or elements of the original study design that preclude the use of the dataset for use in the FSQV reliability assessment.

2.4.1 MINIMAL DATA QUALITY REQUIREMENTS

The use of historical data from multiple studies and locations has inherent problems with comparability and usability. Each study was originally designed to meet the objectives of the investigators and study sponsors. The lack of regionally promulgated guidance for conducting freshwater sediment evaluations contributes to the variability between studies. Therefore, in order to evaluate the relative completeness and subjective quality of a dataset and to determine whether a particular data set was acceptable for use in deriving SQV's, a set of minimal data requirements were established:

1) Synoptic chemistry and bioassay data;
2) Data complete and readily available;
3) Study conducted after 1990;
4) Accepted analytical methods and laboratory QA (SW-846, PSEP, or equivalent);
5) Standard bioassay protocols and QA/QC controls (ASTM procedures or equivalent).

Datasets were considered acceptable if they meet all five of the minimal data requirements, unless a clear fatal flaw is identified that rendered the data set unusable (Section 2.5.2). The primary objective of this method was to maximize the available data for use while maintaining enough commonality to allow for a meaningful comparison between datasets.

Information regarding 14 of the studies reviewed was obtained from data previously entered into SEDQUAL. No additional documentation was available for review of these 14 studies, including confirmation of analytical and biological testing methods. It was generally assumed, based on the dates (post-1993) and sources of these studies (environmental consultants and regulatory agencies), that current methods were employed.

2.4.2 ADDITIONAL QA CONSIDERATIONS

The second, more subjective component of the dataset screening evaluation was the additional QA considerations. This component is more difficult to characterize, in the sense that it takes into account study-specific issues and QA problems. Additional QA considerations that were identified included analytical problems noted during data validation, elevated detection limits, exceeded holding times, limited bioassay replication, and inadequate QA documentation. Information regarding 14 of the studies reviewed was obtained from data previously entered into SEDQUAL. No additional documentation was available for review of these 14 studies, thereby limiting the assessment of potential QA problems usually discussed in data validation reports or QA summaries.

2.5 Results

A total of 58 datasets were reviewed, checklists completed, and screening evaluations performed. Thirty-nine of the datasets evaluated were deemed acceptable for use in the reliability assessment. The remaining 19 studies were rejected for failing to meet one or more of the minimal data quality requirements or due to other problematic QA considerations. The summary of results of the screening evaluation are presented in Table 2.1.

2.5.1 Acceptable Data Sets

A total of 39 datsets met the minimal data quality requirements during the screening evaluation. A summary of the data types for the datasets determined to be acceptable for FSQV development and the reliability assessment is presented in Table 2.2. The majority of the accepted datasets included sediment chemistry results for multiple chemical groups. The most frequently conducted bioassays included *C. tentans* survival and growth, *H. azteca* survival, and Microtox bioluminescence.

Potential limitations to the overall data quality, despite meeting minimal requirements, include: unvalidated results, unavailable information regarding QA/QC measures, uncertainty regarding analytical methods (for studies reviewed from SEDQUAL), low bioassay replication (< 5 replicates), and unreported problems with bioassay testing (i.e. QA/QC, water quality, control performance).

2.5.2 Rejected Data Sets

A total of 19 datasets reviewed were deemed unacceptable for use in FSQV development and the reliability assessment. The major reasons for rejecting datasets based on failing to meet minimal data quality requirements, in order of frequency (note: most rejected studies missed more than one data quality requirement) included: 1) accepted analytical methods and laboratory QA (10 datasets); 2) standard bioassay protocols with QA/QC controls (10 datasets); 3) date of study prior to 1990 (9 datasets); 4) data complete and readily available (9 datasets); and 5) synoptic chemistry and bioassay data (4 datasets). Additional considerations noted among the rejected datasets included a general lack of any available QA information, low bioassay replication, exceeded holding times, and elevated detection limits.

Table 2-1: Qualitative Screening of Freshwater Data Sets

Study Information				Minimal Data Quality Requirements: Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
Dataset (Study Name)	Citation	Study Year	SEDQUAL Survey ID or Other Electronic Data Format	Synoptic chemistry and bioassay data	Data complete and readily available	Study Conducted after 1990[1]	Accepted Analytical Methods and Laboratory QA[2]	Standard Bioassay Protocols with QA/QC Controls[3]	Major QA Problems Identified?	Comments	Does dataset meets minimal requirements for use in reliability assessment?
Gas Works Park, Lake Union	York, Norton, and Stinson (1986)	1985	GWPLKUN	Yes	Yes	No	No	Yes	No	Incomplete QA information	No
Screening Survey for Chemical Contaminants and Toxicity in Drainage Basins at Paine Field	Johnson and Norton (1988)	1987	PAINEFLD	Yes	No	No	Yes	Yes	No	No sampling location information, organic analytes generally exceeded recommended detection limits, low bioassay replication, no positive control or reference.	No
Contaminants in 5 Lower Columbia Ports	Ecology (1988)	1987	LWRCOLUM	Yes	Yes	No	?	No	No	Information collected from SEDQUAL - No reports or other documentation available for review; 2 replicate bioassays	No
Kalama Chemical, Inc., Columbia River	Heffner (1989)	1988	KALAMA88	Yes	Yes	No	No	No	Yes	Incomplete QA information; inadequate replication, water quality problems	No
Ferndale Wastewater Treatment Plant, Wa	Ruiz (1989)	1988	FERNDALE	Yes	No	No	Yes	No	Yes	No holding time or QA/QC information, no bioassay laboratory controls indicated, screening indicated D. magna data but no values entered into SEDQUAL.	No

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
An Assessment of Metals Contamination in Lk. Roosevelt, Wa	Johnson, Norton, and Yake (1989)	1988	Not Known	Yes	Yes	No	No	Yes	Yes	Metals only; no pos. or neg controls run on bioassays, problems noted with reference material, and QA report says that bioassay results should not be considered conclusive. The data may also be included in the Johnson (1991) study , listed as SEDQUAL ID: LAKEROOS	No
Pacific Wood Treating Corporation Class II Inspection, Ridgeway WA	Reif (1989)	1989	PWTC2	Yes	No	No	No	No	Yes	Uncertain data collection protocols, some holding times exceeded but most were not even reported, inadequate bioassay replication, insufficient information for data validation or determining accuracy, icrotox data not entered in SEDQUAL.	No
Report of Findings Vancouver WA Phase II Columbia River	CWEC (1989)	1989	Not Known	No	No	No	No	No	No	No QA/QC information provided, Copper was only metal analyzed, Acute static bioassay used fish	No
Review of Metals, Bioassay, and Macroinvertebrate Data from Lake Roosevelt	Johnson (1991)	1989	LAKEROOS	Yes	Yes	No	Yes	Yes	No	Detection limits and holding times were not identified and it is unclear if pos. controls and references were run for bioassays. QA was not included.	No
Port of Vancouver Results of Daphnia magna Sediment Bioassays	Ecology (1990)	1990	POV89_EI	No	No	Yes	Yes	No	No	Missing data, no QA/QC information provided, copper was only metal analyzed; limited bioassay replication (2)	No
Weyerhaeuser, Longview Pulp & Paper Mill, WA	Andreasson (1991)	1990	WEYLONG	Yes	Yes	Yes	Yes	Yes	No		Yes
Longview Fiber Company, Class II Inspection, Longview WA	Das (1991)	1990	LONGVW90	Yes	Yes	Yes	No	No	Yes	No information on laboratory controls, Some holding times exceeded, screening indicated *D. magna* data but no values entered in SEDQUAL.	No
Reynolds Metal Company Class II Inspection	Heffner (1991)	1990	REYNOLDS	Yes	Yes	Yes	Yes	Yes	Yes	No QA informatioon, bioassay quality uncertain, poor detection levels and sample preparation for metals.	No

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
Alcoa Class II Inspection, Vancouver WA	Zinner (1990)	1990	ALCOA90	Yes	No	Yes	Yes	No	No	No QA information, incomplete data, and limited bioassay information (no holding times or methodology).	No
Steilacoom Lake Sediments	Bennett and Cubbage (1992a)	1990	STEILLK2	Yes	Yes	Yes	Yes	Yes	No	Cannot verify holding times	Yes
McCormick & Baxter Remedial Investigation, Phases 1 & 2	PTI (1992)	1990	MBCREOS1, MBCREOS2	Yes	Yes	Yes	Yes	Yes	No		Yes
PAH's in Lk. Wa at Quendall-Baxter Phase I	Norton (1991)	1990	QUEBAX1	Yes	Yes	Yes	Yes	Yes	No		Yes
Marco Shipyard	Friedman and Bruya (1990)	1990	MARCO90	Yes	Yes	Yes	Yes	Yes	No	SVOC detection limits are above target limits; minimal information available on bioassay quality	Yes
Columbia Slough Sediment Analysis and Remediation Phase I Vol. I and Vol II, North Portion of Portland	Dames & Moore (1991)	1991	CBSLOUGH	Yes	Yes	Yes	Yes	Yes	No	Incomplete report reviewed and some missing information.	Yes
Unimar Drydock, Lake Union, Seattle, WA	FishPro and GeoEngineers (1991)	1991	UNIMAR2	Yes	Yes	Yes	Yes	Yes	No		Yes
Site Hazard Assessment Report, Hansville Landfill, Kitsap County, WA	SAIC (1991)	1991	HANSVL91	Yes	Yes	Yes	Yes	Yes	No		Yes
Effect of PAHs in Sediments form Lake Washington on Freshwater Bioassay Organisms and Benthic Macroinvertebrates	Bennett and Cubbage (1992b)	1991	QUEBAX2	Yes	Yes	Yes	Yes	Yes	No		Yes
Results of Sediment Sampling in Baxter Cove, Lake Washington	Norton (1992)	1991	QUEBAX3	Yes	Yes	Yes	Yes	Yes	No	Holding times not included; no information regarding bioassay QA procedures	Yes

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
Determination of Miscellaneous Metals…from Milltown Resevoir and Clark Fork River, Montana	USGS (1993)	1991	Not Known	Yes	No	Yes	No	No	No	Four replicate bioassays	No
Sediment Sampling and Analysis Report, Cedar River Sediments, Renton WA	Golder Associates (1992)	1992	CEDARIV	Yes	Yes	Yes	Yes	Yes	No		Yes
Mill Creek and East Drain Sediment Sampling Report, Western Processing Phase II, Kent, WA	Landau Associates (1993)	1992	MILLCRP2	Yes	No	Yes	No	Yes	Yes	Analytical methods and QA not adequately discussed, questionable sampling location data, organics data not acceptable, *C. tentans* methods non-standard	No
Lake Union Dry Dock	Hart Crowser (1992)	1992	LKUNDRDK	Yes	Yes	Yes	Yes	Yes	No		Yes
Sediment Quality Assessment of Franklin D. Roosevelt Lake and the Upstream Reach of the Columbia River, Wa	USGS (1992)	1992	LAKROO92	Yes	Yes	Yes	Yes	Yes	No	No pos. control run for bioassay.	Yes
Survey of Contamination in Sediments in Lake Union and Adjoining Waters	Cubbage (1992)	1992	LKUNION	Yes	Yes	Yes	Yes	Yes	No	Detection limits and holding times for sediment chemistry were above acceptable limits.	Yes
Results of Acute Toxicity Tests on Freshwater Sediments Collected from Silver Lake WA using Hyallela azteca and Microtox	Parametrix (1993)	1992	Not Known	No	Yes	Yes	Yes	Yes	No	Data are not synoptic - they were collected several months apart	No
Class II Inspection Boise Cascade, Wallula, WA	Johnson and Heffner (1993)	1993	BOISECAS	Yes	Yes	Yes	Yes	Yes	No	Data acceptable for use as qualified	Yes

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements: Phase 1 Screening Criteria A	B	C	D	E	Additional Considerations		Determination
Columbia Aluminum Company Baseline Sediment Characterization	ENSR 1994b	1993	COLALU94	Yes	Yes	Yes	Yes	Yes	No		Yes
Reconnaissance Investigation of Water Quality, Bottom Sediment, and Biota Associate with Irrigation Drainage in the Columbia Basin Project, 1991-1992;	Embry and Block (1995)	1993	COLBSN92	Yes	Yes	Yes	No	No	No	Holding times were not listed and QA information not included.	No
Alcoa Vancouver Works: Baseline Sediment Characterization	ENSR 1994a	1993	VALCOA93	Yes	Yes	Yes	Yes	Yes	No	QA report not included.	Yes
Lower Columbia Backwater Recon. Survey	Tetra Tech (1993)	1993	LCBWRS93	Yes	Yes	Yes	Yes	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review	Yes
Seattle Commons Parcel C	Shannon and Wilson (1994)	1994	SEACOM94	Yes	Yes	Yes	Yes	Yes	No		Yes
Everett Simpson Site Sediment Investigat	Ecology (1994a)	1994	EVRTSM94	No	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used; small COC list; compositing precludes synoptic dataset	No
Spokane River PCB Bioassay Study	Ecology (1994b)	1994	SPOKNR94	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
Salmon Bay Study Phase III	Serdar, Cubbage, and Rogowski (2000)	1997	SALIII97	Yes	Yes	Yes	Yes	Yes	Yes	TBT and PCBs data should not be used due to analytical problems that resulted in uncertainty regarding the accuracy of the results	Yes
Tri-Star Marine	Ecology (1997)	1997	TRI-STAR	Yes	Yes	Yes	Yes	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review	Yes
Portland Shipyard Sediment Investigation	SEA (1998)	1998	PSYSEA98	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes
Portland Shipyard Environmental Audit, Cascade General	Dames & Moore (1997)	1998	PSYD&M97	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes
Willamette River data (Portland Shipyard Env. Audit)	Dames & Moore (1998)	1998	WRD&M98	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes
Port of Portland, Sediment Characterization Study, Terminals 2 (Berths 203-206) and Marine Terminal 4 (Berth 416)	Degens (1998)	1998	PPTLDT24	Yes	Yes	Yes	Yes	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review	Yes
Terminal 4 Remedial Investigation (Port of Portland)	Quinn (1998)	1998	WLRPT498	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
Tosco Dredged Material Evaluation (TOSCO)	TOSCO (1997)	1999	TOSCO99	Yes	Yes	Yes	Yes	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review	Yes
Lake Sammamish Baseline Sediment Stdy 99	King County (1999)	1999	LSAMM99	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; acceptance assumes standard analytical procures were used	Yes
Port of Portland Site Investigation Report, Ross Island	Hart Crowser (2000)	2000	ROSSIS99	Yes	Yes	Yes	Yes	Yes	No		Yes
Chemical Analysis and Toxicity Testing of Spokane River Sediments	Johnson and Norton (2001)	2000	SPOK2000	Yes	Yes	Yes	Yes	Yes	No		Yes
Review of Sediment Quality Data for Similkameen River	Ecology (2000)	2000	SIMILK00	Yes	Yes	Yes	Yes	Yes	No		Yes
Lake Washington Baseline Sed Study 2000	King County (2000a)	2000	LKWA00	Yes	Yes	Yes	?	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review	Yes
Lake Union University Regulator CSO Post Separation Study 2000	King County (2000b)	2000	LUUCSO00	Yes	Yes	Yes	?	Yes	Yes	Information collected from SEDQUAL - No reports or other documention available for review; Bioassays exceeded holding times	Yes
Sediment Sampling for Quendall Terminals Property	Exponent (2001)	2001	QUEDAL00	Yes	No	Yes	Yes	Yes	No	Limited sediment chemistry	Yes
Cargill Irving Elevator Terminal	Harding ESE (2001)	2001	CARGIL01	Yes	Yes	Yes	Yes	Yes	Yes	TBT data and chironomus growth results are useable; however, results should be used with caution and considered as estimates. Minor holding time exceedances.	Yes

Table 2-1: Qualitative Screening of Freshwater Data Sets (Continued)

Study Information				Minimal Data Quality Requirements – Phase 1 Screening Criteria					Additional Considerations		Determination
				A	B	C	D	E			
McCormick and Baxter Creosoting Company Sediment Remedial Design	Ecology and Environment (2001)	2001	MBCREOS3, MBCREOS4	Yes	Yes	Yes	Yes	Yes	No	Bioassay samples exceeded storage temperatures when received by lab; *H. azteca* re-tested due to control failure; re-test data are acceptable	Yes
Reassessment of Toxicity of Lk. Roosevelt Sediments, Nov 2001 Draft	Ecology (2001)	2001	LKROOS01	Yes	Yes	Yes	Yes	Yes	No	Metals only.	Yes
BNSF Skykomish River Site	BNSF (2002)	2001	BNSFSK02	Yes	Yes	Yes	Yes	Yes	No	Information collected from SEDQUAL - No reports or other documention available for review; Hg only analyte, concerns regarding bioassay QA	No
Lower Willamette River Reference Area Study	Hart Crowser (2002)	2001	Not Entered	Yes	Yes	Yes	Yes	Yes	No		Yes

Notes:

1: Pre-1990 datsets will be considered if acceptable methods were utilized and full QA back-up information is available. A notation will be made in the commnets section for any datasets meeting this caveat.

2: SW-846, PSEP protocols or equivalent; information regarding analytical methods was not available for some studies reviewed using SEDQUAL data entries--for these datasets, their acceptance is conditional on the assumption that standard methods were used due to their relevant recency and project sponsors (environmental consultants and regulatory agencies).

3: ASTM procedures or equivalent

Number of data sets screened: 58

Number of data sets rejected: 19

Number of data sets accepted: 39

Table 2-2: Data Summary for Accepted Datasets

					Chemical Groups						Bioassays[4]								
Study Name	Citation[1]	Survey Date[2]	Waterbody	SEDQUAL Survey ID or Other Electronic Data Format	TOC	Other Conventional	Metals	SVOCs[3]	Pest/ PCB	PAHs	*Chironomus tentans*	*Hyallela azteca*	*Daphnia magna*	Microtox	Other Bioassay	Benthic Community[5]	Data Validation[6]	QA Data Available[7]	Original Data Reports Reviewed[8]	Comments/ Problems Noted
McCormick & Baxter Remedial Investigation, Phases 1 & 2	PTI (1992)	1990	Willamette River, OR	MBCREOS1, MBCREOS2	Y	Y	Y	Y	Y	Y	0	48	0	0	0	0	QA2	U	Y	5 replicate bioassays
PAH's in Lk. Wa at Quendall-Baxter Phase I	Norton (1991)	1990	Lake Washington, WA	QUEBAX1	N	Y	Y	Y	Y	Y	0	4	4	0	0	Y	U	U	Y	5 replicate bioassays; benthic samples taken but not yet entered in Sedqual;
Marco Shipyard	Friedman and Bruya (1990)	1990	Lake Washington Ship Canal, WA	MARCO90	Y	Y	Y	Y	Y	Y	0	3	0	0	0	0	SR	P	Y	3 replicates for *H. azteca*; SVOC detection limits exceed target detection limits; limited information on bioassay quality
Weyerhaeuser, Longview Pulp & Paper Mill, Wa	Andreasson (1991)	1990	Columbia River, WA	WEYLONG	Y	Y	Y	Y	Y	Y	0	3	0	0	0	0	U	P	Y	'5 replicates for *H. azteca*
Steilacoom Lake Sediments	Bennett and Cubbage (1992)	1990	Steilacoom Lake	STEILLK2	Y	Y	Y	N	N	N	4	4	4	4	4	4	U	P	Y	Holding time information not available; microtox data not entered in SEDQUAL
Columbia Slough Sediment Analysis and Remediation Phase I Vol. I and Vol II, North Portion of Portland	Dames & Moore (1991)	1991	Columbia River, WA and Columbia Slough, OR	CBSLOUGH	Y	Y	Y	Y	Y	Y	0	19	0	0	0	0	SR	P	Y	incomplete report reviewed; some missing data; 5 replicates for *H. azteca*
Effect of PAHs in Sediments form Lake Washington on Freshwater Bioassay Organisms and Benthic Macroinvertebrates	Bennett and Cubbage (1992b)	1991	Lake Washington, WA	QUEBAX2	N	Y	Y	Y	Y	Y	3	3	3	0	3	Y	U	U	Y	Benthic samples taken but not yet entered in Sedqual;
Results of sediment sampling in Baxter Cove, Lake Washington	Norton (1992)	1991	Lake Washington, WA	QUEBAX3	Y	Y	N	Y	N	N	0	3	0	0	0	Y	U	P	Y	Benthic samples taken but not yet entered in Sedqual; no holding time information and limited data results; 5 replicates for *H. azteca*
Unimar Drydock, Lake Union, Seattle, Wa	FishPro and GeoEngineers (1991)	1991	Lake Union, WA	UNIMAR2	N	N	Y	Y	Y	Y	0	9	0	0	0	0	U	P	Y	5 replicate bioassays
Site Hazard Assessment Report, Hansville Landfill, Kitsap County, WA	SAIC (1991)	1991	Not Known	HANSVL91	N	N	Y	Y	Y	N	0	2	0	0	0	0	U	P	Y	5 replicate bioassays

Table 2-2: Data Summary for Accepted Datasets (Continued)

					Chemical Groups						Bioassays[4]								
Study Name	Citation[1]	Survey Date[2]	Waterbody	SEDQUAL Survey ID or Other Electronic Data Format	TOC	Other Conventional	Metals	SVOCs[3]	Pest/ PCB	PAHs	*Chironomus tentans*	*Hyallela azteca*	*Daphnia magna*	Microtox	Other Bioassay	Benthic Community[5]	Data Validation[6]	QA Data Available[7]	Original Data Reports Reviewed[8]	Comments/ Problems Noted
Sediment Sampling and Analysis Report, Cedar River Sediments, Renton Wa	Golder Associates (1992)	1992	Cedar River, Renton WA	CEDARIV	Y	Y	Y	Y	Y	Y	0	5	5	0	0	0	U	P	Y	5 replicates for *H. azteca* and 2 reps. for *C. tentans*
Class II Inspection of the Boise Cascade Pulp and Paper Mill, Wallula WA	Johnson and Heffner (1993)	1992	Lake Wallula, WA	BOISECAS	Y	Y	Y	Y	Y	N	0	5	0	0	0	0	U	P	Y	5 replicate bioassays
Lake Union Dry Dock	Hart Crowser (1992)	1992	Lake Union, WA	LKUNDRDK	Y	Y	Y	Y	Y	Y	0	4	0	0	0	0	QA1	C	Y	5 replicate bioassays
Sediment Quality Assessment of Franklin D. Roosevelt Lake and the Upstream Reach of the Columbia River, Wa	USGS (1992)	1992	Lake Roosevelt, WA	LAKROO92	Y	Y	Y	Y	N	Y	0	22	0	Y	22	0	U	N	Y	3 replicates for *H. azteca*; 10 replicates for *C. dubia;* microtox data not entered into Sedqual
Results of Acute Toxicity Tests on Freshwater Sediments Collected from Silver Lake WA using Hyallela azteca and Microtox	Parametrix (1993)	1992	Silver Lake, WA	Not Known	N	N	Y	Y	Y	N	0	Y	0	Y	Y	?	U	P	Y	Bioassays and chemical data collected several months apart
Survey of Contamination in Sediments in Lake Union and Adjoining Waters	Cubbage (1992)	1992	Lake Union, WA	LKUNION	Y	Y	Y	Y	Y	Y	0	9	9	0	0	Y	SR	P	Y	Detection limits and holding times for chemical analyses exceeded acceptable limits; 5 replicate bioassays
Alcoa Vancouver Works: Baseline Sediment Characterization	ENSR (1994a)	1993	Columbia River, WA	VALCOA93	Y	Y	Y	N	Y	Y	0	4	0	0	0	0	QA1	C	Y	5 replicate bioassays
Columbia Aluminum Company Baseline Sediment Characterization	ENSR (1994b)	1993	Columbia River, WA	COLALU94	Y	Y	Y	Y	N	Y	0	6	0	6	0	0	QA2	C	Y	High variability (survival) in bioassay results despite acceptable test parameters; 5 replicate bioassays
Lower Columbia Backwater Reconnaisance Survey	Tetra Tech (1993)	1993	Columbia River, WA	LCBWRS93	Y	Y	Y	Y	Y	Y	0	15	0	0	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed;
Seattle Commons Parcel C	Shannon and Wilson (1994)	1994	Lake Union, WA	SEACOM94	Y	Y	Y	Y	Y	Y	0	3	0	3	0	0	U	P	Y	5 replicate bioassays

Table 2-2: Data Summary for Accepted Datasets (Continued)

					Chemical Groups						Bioassays[4]								
Study Name	Citation[1]	Survey Date[2]	Waterbody	SEDQUAL Survey ID or Other Electronic Data Format	TOC	Other Conventional	Metals	SVOCs[3]	Pest/ PCB	PAHs	*Chironomus tentans*	*Hyallela azteca*	*Daphnia magna*	Microtox	Other Bioassay	Benthic Community[5]	Data Validation[6]	QA Data Available[7]	Original Data Reports Reviewed[8]	Comments/ Problems Noted
Everett Simpson Site Sediment Investigation	Ecology (1994a)	1994	Columbia River, WA	EVRTSM94	N	N	Y	N	N	N	4	4	0	0	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed; only mercury analyzed
Spokane River PCB Bioassay Study	Ecology (1994b)	1994	Spokane River, WA	SPOKNR94	Y	Y	Y	Y	Y	Y	0	3	0	3	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed;
Tri-Star Marine	Ecology (1997)	1997	Lake Washington Ship Canal, WA	TRI-STAR	N	Y	Y	Y	Y	Y	0	3	0	3	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed; 5 replicates for H. azteca and 5 replicates for Microtox
Salmon Bay Study Phase III	Serdar, Cubbage, and Rogowski (2000)	1997	Lake Washington Ship Canal, WA	SALIII97	Y	Y	Y	Y	Y	Y	22	22	0	22	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed; 5 replicates for C. tentans, H. azteca and Microtox
Portland Shipyard Sediment Investigation	SEA (1998)	1998	Willamette River, OR	PSYSEA98	Y	Y	Y	Y	Y	Y	55	55	0	55	0	0	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed; 8 replicates for *H. azteca* and *C. tentans*; 5 replicates for Microtox
Portland Shipyard Environmental Audit, Cascade General	Dames & Moore (1997)	1998	Willamette River, OR	PSYD&M97	Y	Y	Y	Y	Y	Y	0	3	0	0	0	5	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed; 5 replicate bioassays
Willamette River data;Portland Shipyard Environmental Audit	Dames & Moore (1998)	1998	Willamette River, OR	WRD&M98	Y	Y	Y	Y	Y	Y	0	2	0	0	0	9	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed; 5 replicate bioassays
Port of Portland, Sediment Characterization Study, Terminals 2 (Berths 203-206) and 4 (Berth 416)	Degens (1998)	1998	Willamette River, OR	PPTLDT24	Y	Y	Y	Y	Y	Y	4	4	0	0	0	0	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed; 8 replicate bioassays

Table 2-2: Data Summary for Accepted Datasets (Continued)

					Chemical Groups						Bioassays[4]								
Study Name	Citation[1]	Survey Date[2]	Waterbody	SEDQUAL Survey ID or Other Electronic Data Format	TOC	Other Conventional	Metals	SVOCs[3]	Pest/ PCB	PAHs	*Chironomus tentans*	*Hyallela azteca*	*Daphnia magna*	Microtox	Other Bioassay	Benthic Community[5]	Data Validation[6]	QA Data Available[7]	Original Data Reports Reviewed[8]	Comments/ Problems Noted
Terminal 4, Slip 3, Sediment Investigation (Port of Portland)	Quinn (1998)	1998	Willamette River, OR	WLRPT498	Y	Y	Y	Y	N	Y	22	22	0	0	0	0	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed;
Tosco Dredged Material Evaluation (TOSCO)	TOSCO (1997)	1999	Willamette River, OR	TOSCO99	Y	Y	Y	Y	Y	Y	3	3	0	0	0	0	QA2	U	N	Information collected from SEDQUAL-no other documentation reviewed; 8 replicate bioassays
Lake Sammamish Baseline Sediment Study	King County (1999)	1999	Lake Sammamish, WA	LSAMM99	Y	Y	Y	Y	Y	Y	17	17	0	17	0	0	QA1	U	N	Information collected from SEDQUAL-no other documentation reviewed;
Chemical Analysis and Toxicity Testing of Spokane River Sediments	Johnson and Norton (2001)	2000	Spokane River, WA	SPOK2000	Y	Y	Y	Y	Y	Y	8	8	0	8	0	0	U	P	Y	8 replicates for *C. tentans* and *H. azteca*; 5 replicates for Microtox
Port of Portland Site Investigation Report (Ross Island)	Hart Crowser (2000)	2000	Port of Portland	Not Entered	N	Y	N	Y	Y	Y	14	14	0	0	0	0	QA1	U	Y	Independent review mentioned but not included with draft report
Review of Sediment Quality Data for Similkameen River	Ecology (2000)	2000	Similkameen River	SIMILK00	Y	Y	Y	N	Y	N	0	4	0	4	0	5	U	P	Y	Bioassay controls and holding times not available; 4 replicates for *H. axteca*; 5 replicates for microtox
Lake Washington Baseline Seidment Study	King County (2000a)	2000	Lake Washington, WA	LKWA00	Y	Y	Y	Y	Y	Y	28	28	0	27	0	0	QA1	U	N	Information collected from SEDQUAL-no other documentation reviewed;
Lake Union University Regulator CSO Post Separation Study	King County (2000b)	2000	Lake Union, WA	LUUCSO00	N	Y	N	N	N	N	7	7	0	7	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed; only conventional parameters entered in SEDQUAL
Reassessment of Toxicity of Lk. Roosevelt Sediments, Nov 2001 Draft	Ecology (2001)	2001	Lake Roosevelt, WA	LKROOS01	Y	Y	Y	N	N	N	10	10	0	10	0	0	U	P	Y	8 replicates for *C. tentans* and *H. azteca*; 5 replicates for Microtox
Sediment Sampling for Quendall Terminals Property	Exponent (2001)	2001	Lake Washington, WA	QUEDAL00	Y	Y	N	N	N	N	9	9	0	9	0	0	U	P	Y	Limited sediment chemistry; 8 replicates for *C.tentans & H. axteca*; 5 replicates for microtox

Table 2-2: Data Summary for Accepted Datasets (Continued)

					Chemical Groups						Bioassays[4]								
Study Name	Citation[1]	Survey Date[2]	Waterbody	SEDQUAL Survey ID or Other Electronic Data Format	TOC	Other Conventional	Metals	SVOCs[3]	Pest/ PCB	PAHs	*Chironomus tentans*	*Hyallela azteca*	*Daphnia magna*	Microtox	Other Bioassay	Benthic Community[5]	Data Validation[6]	QA Data Available[7]	Original Data Reports Reviewed[8]	Comments/ Problems Noted
McCormick and Baxter Creosoting Company Sediment Remedial Design	Ecology and Environment (2001)	2001	Willamette River, OR	MBCREOS3, MBCREOS4	Y	Y	Y	N	N	Y	61	61	0	0	0	0	QA1	C	Y	Bioassay samples exceeded storage temperatures when received by lab
Cargill Irving Elevator Terminal	Harding ESE (2001)	2001	Willamette River, OR	CARGIL01	N	Y	Y	N	Y	Y	3	3	0	0	0	0	SR	P	Y	TBT data and *C. tentans* growth results not useable; minor holding time exceedances
BNSF Skykomish River Site	BNSF (2002)	2001	Skykomish River, WA	BNSFSK02	Y	Y	N	N	N	N	6	6	0	6	0	0	U	U	N	Information collected from SEDQUAL-no other documentation reviewed; only conventional parameters entered in SEDQUAL
Lower Willamette River Reference Area Study	Hart Crowser (2002)	2001	Willamette River, OR	Not Entered	Y	Y	Y	Y	Y	Y	3	3	0	0	3	0	QA1	C	Y	Bioaccumulation testing was also conducted using *Lumbriculus variegatus* and *Corbicula fluminea*

3.0 RELIABILITY ASSESSMENT OF SQV SETS (TASK 5)

3.1 Introduction

This section provides a reliability assessment of existing freshwater sediment quality value (SQV) sets in North America, as a step toward identifying or developing freshwater SQVs for use in Washington State. This report presents the results of Task 5. The goals of this task were to:

- Conduct a reliability assessment of eight existing freshwater SQV sets selected in Task 2, against a regional synoptic freshwater data set assembled in Tasks 3 and 4. The reliability assessment provided an assessment of the false negative rates, false positive rates, and overall reliability of each candidate SQV set.

- Determine whether any of the existing SQV sets has adequate reliability for use in regulatory programs in Washington State. If so, make recommendations regarding which SQV sets are appropriate for which purposes (e.g., various levels of protectiveness).

- If none of the existing SQV sets had adequate reliability, determine whether updated freshwater AETs or alternative SQV sets can be developed as part of Phase II, and make recommendations regarding data collection and/or calculation techniques that could be employed.

The steps involved in conducting the reliability assessment include the following:

- **Data Preparation –** The database was examined to ensure that all data were reported in appropriate units and corrections were made where needed. Individual samples and stations without synoptic data were removed from the data set.

- **Chemical List Finalization –** The list of chemicals to be included in the reliability assessment was refined to include any chemical appearing on at least one SQV list for which there was at least one detected value in the database.

- **Bioassay and Endpoint Selection –** The final list of bioassays and endpoints used in the reliability analysis was developed based on the data available, and the freshwater bioassays and endpoints for which standard methods are available.

- **Biological Hit/No-Hit Definitions –** For each bioassay and endpoint selected, biological hit/no-hit definitions were developed at three conceptual levels – statistically significant difference, SQS, and CSL. While these terms are used for convenience, they are intended to represent more generally a no adverse effects level, a level above which minor adverse effects may occur, and a level above which more significant adverse effects may occur, as used in any of the regional sediment management programs (dredging, source control, and cleanup). The individual bioassay/endpoint hit/no-hit definitions were then combined into overall hit/no-hit definitions for individual stations. These definitions were used to generate biological hit/no-hit lists for the stations at each of the three effects levels.

- **Reliability Assessment –** An Excel spreadsheet with Visual Basic macros was used to compare the chemistry data with the numeric guidelines in the SQV sets to determine chemical hit/no-hit predictions for each station. These predictions were compared to the biological hit/no-hit lists to calculate the various estimates of reliability for each effects level and SQV set combination. In addition, three methods of data interpretation were tested: 1) comparison to control, 2) comparison to

reference, defaulting to control if a valid reference was not available, and 3) comparison to reference only, eliminating any stations that did not have a valid reference. This combination of three data comparison methods with three effects levels resulted in nine reliability runs for each of the eight SQV sets.

- **Alternatives Assessment –** A brief evaluation of alternative SQV sets (e.g., updated AETs and floating percentile values) was conducted using the spreadsheet to evaluate the likelihood that Phase II efforts would result in SQV sets with significantly better reliability than existing SQV sets.

Each of these steps is described in greater detail below.

3.2 Methods

This section describes the methods that were used to prepare for and conduct the reliability assessment, beginning with the database assembled in Tasks 3 and 4, and the eight SQV sets selected for assessment in Task 2.

3.2.1 DATA PREPARATION

All of the freshwater data sets remaining after QA screening in Task 4 were assembled in a project database consisting of only these surveys within a SEDQUAL Information System Version 4.2 shell. The following data preparation steps were carried out:

- Many surveys contained stations or individual samples that had only chemistry data but no bioassay data, or only bioassay data with no chemistry data. These samples and stations are not useful for the reliability analysis and were deleted. In addition, there were some surveys and samples with only conventionals data; these samples were also deleted.

- Data for bioassays not being used in the reliability analysis were deleted. Some bioassay tests were eliminated because there were not enough data for that test, and some individual data points were eliminated due to quality assurance issues, such as not having enough replicates or overluminescence in the Microtox bioassay.

- Chemistry data were examined to ensure that the data had the correct number of significant digits and that units were properly expressed and consistent throughout the data set. Where errors were found, chemistry templates were exported, corrected, and re-imported.

- Bioassay data were examined to ensure that all data were entered correctly with respect to bioassay variable, species and NODC codes, and bioassay units. These all affect the ability of the bioassay statistical analysis (BSA) tool in the SEDQUAL Information System to recognize these data sets and process them correctly. Where errors or inconsistencies were found, bioassay data were exported, corrected, re-imported, and double-checked to ensure that the BSA tool was processing each survey correctly.

- The resulting database was used for the runs in which samples were compared to control and to mixed reference and control. A subset of this database was prepared and used for the runs in which samples were compared to reference only, consisting only of those data that had valid reference samples. Samples were excluded if there were no reference samples in the data set, or if the reference samples failed quality assurance review.

3.2.2 CHEMICAL LIST FINALIZATION

The chemistry data set was examined and compared to the list of chemicals included in the eight SQV sets selected for reliability analysis. Any chemicals for which no data are available in the database or for which all data are undetected values were removed from the reliability assessment SQV sets and spreadsheets. The remaining chemicals are all chemicals which are included in at least one SQV set and which have at least one detected value in the database. These chemicals are listed below:

- **Metals:** Antimony, arsenic, cadmium, chromium, copper, lead, mercury, nickel, silver, zinc.
- **Conventionals:** Ammonia, sulfides, TOC.
- **PAHs:** Naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benz(a)anthracene, chrysene, benzofluoranthenes, benzo(a)pyrene, indeno(1,2,3-cd)pyrene, dibenz(ah)anthracene, benzo(ghi)pyrelene, total LPAH, total HPAH.
- **Other Organics:** 2,3,7,8-TCDD, bis(2-ethylhexyl)phthalate, carbazole, dibenzofuran, di-n-butyl phthalate, phenol, Aldrin, Dieldrin, Endrin, gamma-BHC, hexachlorobenzene, DDT, Aroclor 1248, Aroclor 1254, Aroclor 1260, total PCBs.

Mercury is reported as elemental mercury and chromium as total chromium, as there is not adequate information in the database to look at methylmercury or Cr^{+6}. The lack of speciation information could contribute to observed variability in the results and reduce overall reliability for all SQV sets. Similarly, in freshwater environments, parameters such as alkalinity, hardness, and pH can affect the bioavailability of metals and ionic organic compounds. These parameters are also not routinely available with most data sets, and therefore their effect on variability and reliability cannot be assessed, though it may be significant.

3.2.3 BIOASSAY AND ENDPOINT SELECTION

After initial data screening, the amount of available freshwater data that remained for each biological test is listed below:

- ***Hyalella azteca*:** 382 stations of acute mortality data, 7 stations of chronic growth and mortality data
- ***Chironomus tentans*:** 199 stations of acute growth and mortality data, 15 stations of chronic data
- **Microtox®:** 199 stations (porewater and deionized water extract)
- ***Daphnia magna*:** 27 stations
- ***Ceriodaphnia dubia*:** 30 stations
- **Other bioassays:** less than 10 stations
- **Benthic data:** 25 stations (not entered into the SEDQUAL Information System)

Based on this dataset, the three most widely-used bioassays (*Hyalella*, *Chironomus*, and Microtox®) were selected for the reliability assessment, as these are also the most likely to be incorporated into regional testing protocols and adopted as part of state sediment programs. These are the same tests preliminarily recommended by Ecology's 1999 Freshwater Sediment Workgroup as the primary acute and chronic tests for regulatory development. All of the other biological tests have 30 or fewer data points, providing very limited data with which to answer key questions related to appropriate endpoints and usability of the test protocol. Bioassays with more than 100 stations of data can be analyzed to assess factors such as natural

variability in test responses, which will aid in establishing test endpoints. In addition, ASTM protocols are available for the *Hyalella azteca* and *Chironomus tentans* acute and chronic tests, and round robin studies have been conducted for them (ASTM 2000). Finally, in various sensitivity tests referenced in the acute ASTM protocols, the *Hyalella* and *Chironomus* tests were found to be more sensitive than the *Daphnia* or *Ceriodaphnia* tests, as well as other possible tests, including *Lumbriculus variegatus*.. Therefore, the following tests and endpoints were selected for development of biological endpoints and inclusion in the reliability testing:

- *Hyalella azteca*: 10-day mortality, 28-day growth and mortality
- *Chironomus tentans*: 10-day growth and mortality, 20-day growth and mortality
- Microtox®: 15-minute reduction in bioluminescence

Historical data sets include tests that are similar but not identical to the above tests. For example, *Hyalella* acute tests have been run at various durations, such as 14-day or 7-day instead of 10-day. These test durations will be treated the same for purposes of the reliability assessment. Similarly, several different versions of Microtox® have been run over the years, including solid phase, porewater, and deionized water extract tests. For the purposes of the reliability assessment, these are treated similarly.

The following draft endpoints for freshwater bioassays were developed based on a variety of sources, including the preliminary results of Ecology's freshwater sediment standards working group, draft SMS freshwater rule language, Ecology's draft final freshwater Microtox® protocol, and information contained in ASTM Standard E1706-00 for acute and chronic *Hyalella* and *Chironomus* methods, including results of round-robin testing, minimum detectable differences, and natural variability of these species in the tests. However, it should be noted that there is no other jurisdiction that has promulgated or developed guidance on appropriate biological endpoints for these tests, other than statistical difference from a reference or control sample.

Both SQS- and CSL-level endpoints are provided below, since both are needed for use in Ecology's regulatory programs. As noted in the introduction, these levels are referred to as SQS and CSL in this report, but are intended to also represent equivalent levels used in regional dredged material management and source control programs. The combination of these various endpoints into hit/no-hit lists for use in the reliability analysis is discussed in Section 3.3. In each case, the phrase "statistically significant" means a statistical difference from a reference sample (or control sample, if there is no reference sample) at an alpha level of 0.05. Selection of reference stations, data transformations, and statistical testing procedures are identical to those currently in use by Ecology and DMMP programs for marine sediment data (Michelsen and Shaw 1996, Fox et al. 1998). The SQS and CSL endpoints are summarized in Table 3-1, along with their associated control and reference performance standards.

Table 3-1. SQS and CSL Endpoints for Biological Tests

Test	QA Control	QA Reference	SQS	CSL
Hyalella azteca 10-day mortality	C ≤ 20%	R ≤ 25%	T – R > 10%	T – R > 25%
Hyalella azteca 28-day mortality	C ≤ 20%	R ≤ 30%	T – R > 10%	T – R > 25%
Hyalella azteca 28-day growth	CF ≥ 0.15 mg/ind	RF ≥ 0.15 mg/ind	T/R < 0.75	T/R < 0.6
Chironomus tentans 10-day mortality	C ≤ 30%	R ≤ 30%	T – R > 10%	T – R > 25%
Chironomus tentans 10-day growth	CF ≥ 0.48 mg/ind	RF/CF ≥ 0.8	T/R < 0.8	T/R < 0.7
Chironomus tentans 20-day mortality	C ≤ 32%	R ≤ 35%	T – R > 15%	T – R > 25%
Chironomus tentans 20-day growth	CF ≥ 0.48 mg/ind	RF/CF ≥ 0.8	T/R < 0.75	T/R < 0.6
Microtox® decrease in luminescence	CF/CI ≥ 0.72	RF/CF ≥ 0.8	T/R < 0.85	T/R < 0.75

C = Control, CI = Control Initial, CF = Control Final
R = Reference, RF = Reference Final
T = Test Sample

***Hyalella azteca* 10-day mortality bioassay**

- **SQS mortality:** Statistically significant difference from reference, and a relative increase in mortality of > 10% (test – reference > 10%).
- **CSL mortality:** Statistically significant difference from reference, and a relative increase in mortality of > 25% (test – reference > 25%).

The ASTM protocols establish 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%. As such, it is recommended that the reference sample performance standard be established at 25% mortality, just above the control sample performance standard. Given this, the maximum possible mortality that could be observed at the SQS level would be 35%, and would often be less, and the maximum possible mortality that could be observed at the CSL level would be 50%, and would often be less. This SQS level would be very similar in practice to the marine SQS level of 30% absolute mortality, but would allow the reference sample to play a role in identifying hits, which is the case for each of the other biological endpoints.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 5 to 24%, with a mean of 11%. Therefore, a detectable difference could be observed at levels as low as 10-15% mortality, ranging in the worst case up to about 40% mortality, depending on the performance of the control and reference samples, and the degree of variability in the test replicates. In practice these thresholds should be observable nearly all of the time, with the minimum detectable difference at times exceeding the SQS numeric threshold, but not likely exceeding the CSL numeric threshold.

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

- **SQS mortality:** Statistically significant difference from reference, and a relative decrease in mortality of > 10% (test – reference > 10%).
- **CSL mortality:** Statistically significant difference from reference, and a relative increase in mortality of > 25% (test – reference > 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. However, due to somewhat higher mortalities observed in the control samples in the 28-day tests, it is recommended that the reference sample performance standard be initially established at 30% mortality. Given this, the maximum possible mortality that could be observed at the SQS level would be 40%, and would often be less, and the maximum possible mortality that could be observed at the CSL level would be 55%, and would often be less. This approach sets the same policy goals as the acute mortality test, but gives a little more latitude in the reference performance standard for the challenges of running a longer test.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 3 to 28%, with a mean of 8%. Therefore, a detectable difference could be observed at levels as low as 15% mortality, ranging in the worst case up to about 50% mortality, depending on the performance of the control and reference samples, and the degree of variability in the test replicates. In practice these endpoints should be observable most of the time, with the minimum detectable difference at times exceeding the SQS numeric threshold, but not likely exceeding the CSL numeric threshold.

- **SQS growth:** Statistically significant difference from reference, and a relative decrease in weight of > 25% (test/reference < 75%).
- **CSL growth:** Statistically significant difference from reference, and a relative decrease in weight of > 40% (test/reference < 60%).

The SQS and CSL endpoints are based largely on the minimum detectable differences reported in ASTM round robin studies, since little additional information exists on which to base recommendations. The mean minimum detectable difference in weight in round robin studies was approximately 25%, with a range from 16 to 50%. Balancing these considerations are literature studies suggesting that reductions in growth of as little as 20-30% can cause significant reproductive effects and other physiological changes in aquatic species, including *Chironomus tentans* and *Mytilus galloprovincialis* (ASTM 2000, 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 should be observable about half the time, and the numeric level corresponding to the CSL should be observable about 80% of the time.

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

In deciding how to combine these two endpoints for hit/no-hit determination, the ASTM protocols report that the two endpoints often act independently in response to contaminated sediments. In an EPA database comprised of 44 samples, 16% of the sediments reduced both survival and growth, 14% reduced

survival only, and 18% reduced growth only (the rest were no-hits). Therefore, the two endpoints are best used in combination – if either one hits, the station is considered a hit for this bioassay.

***Chironomus tentans* 10-day mortality and growth bioassay**

- **SQS mortality:** Statistically significant difference from reference, and a relative decrease in mortality of > 10% (test – reference > 10%).
- **CSL mortality:** Statistically significant difference from reference, and a relative increase in mortality of > 25% (test – reference > 25%).

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 7%, with a range of 0-15%. As such, it is recommended that the reference sample performance standard also be established at 30% mortality, since the actual control performance in this test is much better than 30%. Given this, the maximum possible mortality that could be observed at the SQS level would be 40%, and would usually be less, and the maximum possible mortality that could be observed at the CSL level would be 55%, and would usually be less.

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

- **SQS growth:** Statistically significant difference from reference, and a relative decrease in weight of > 20% (test/reference < 80%).
- **CSL growth:** Statistically significant difference from reference, and a relative decrease in weight of > 30% (test/reference < 70%).

The SQS and CSL endpoints are based largely on the minimum detectable differences reported in ASTM round robin studies. The mean minimum detectable difference in weight in round robin studies was approximately 11%, with a range from 5 to 24%. This allows for more protective SQS and CSL levels than for either of the chronic growth tests. The round robin studies suggest that the numeric level corresponding to the SQS should be observable well over half of the time, and the CSL levels should be observable nearly all of the time. The numeric levels chosen span the range of growth rates associated with adverse reproductive or physiological effects in the literature, as discussed above.

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

As discussed above for the *Hyalella* 28-day test, it is recommended that if either the growth or mortality levels are exceeded, the station should be considered a hit. In practice, it is likely that the growth endpoint will be the more sensitive of the two.

***Chironomus tentans* 20-day mortality and growth bioassay**

- **SQS mortality:** Statistically significant difference from reference, and a relative decrease in mortality of > 15% (test – reference > 15%).
- **CSL mortality:** Statistically significant difference from reference, and a relative increase in mortality of > 25% (test – reference > 25%).

As noted above, the ASTM protocols establish a control performance standard of 30% mortality for the acute test, which was used as a target for round robin testing of the chronic test. However, the mean mortality observed in the control samples in round robin testing was substantially higher in the chronic test than in the acute test, with 22% of the laboratories failing the acute performance standard. Mean mortality was approximately 25%, with a range of 3-40%. Based on examination of the existing freshwater data set, it was clear that a significantly higher percentage of the data could be retained if the control performance standard was increased from 30% to 32%, and therefore, that was done for the purposes of reliability testing using historic data. As such, it is recommended that the reference sample performance standard be established at 35% mortality. As discussed below, the proposed SQS relative numeric level is increased to 15% for this bioassay because the minimum detectable difference is higher than for other mortality bioassays due to variability among replicates.

Given this, the maximum possible mortality that could be observed at the SQS level would be 50%, and the maximum possible mortality that could be observed at the CSL level would be 60%. Once bioassay procedures and results are stabilized, the reference sample performance standards can be tightened and the maximum mortality at the SQS and CSL levels will decrease.

In ASTM round robin testing, the minimum detectable difference between the test and control sample ranged from 6 to 25%, with a mean of approximately 15%. Therefore, a detectable difference could be observed at levels as low as 10% mortality, ranging in the worst case up to about 60% mortality, depending on the performance of the control and reference samples, and the degree of variability in the test replicates. The SQS numeric thresholds should be observable about half the time, and the CSL thresholds should be observable nearly all of the time.

- **SQS growth:** Statistically significant difference from reference, and a relative decrease in weight of > 25% (test/reference < 75%).
- **CSL growth:** Statistically significant difference from reference, and a relative decrease in weight of > 40% (test/reference < 60%).

The SQS and CSL 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 varied from 25 to 65% for different sediments (the only test endpoint to show significant variation between sediments), ranging from 15 to 125%. The round robin studies suggest that this test is still highly variable, and that levels below those proposed above may not be observable much of the time. The same control and reference performance standards are used as for the 10-day test.

As discussed above, it is recommended that if either the growth or mortality levels are exceeded, the station should be considered a hit, especially since hits may be more difficult to observe in this test due to high variability.

Microtox® luminescence bioassay

- **SQS:** Statistically significant difference from reference, and a relative reduction in luminescence of = 15% (test/reference < 85%).
- **CSL:** Statistically significant difference from reference, and a relative reduction in luminescence of = 25% (test/reference < 75%).

These endpoints are based on Ecology's draft final Microtox® protocol for the revised 100% porewater method (Ecology 2002). Although some of the historic data were not collected using this method, it is recommended that the reliability assessment use the SQS and CSL levels that are expected to be applied in the future, in order to assess which of the existing SQV sets best predicts these levels of effects. Although these levels are inconsistent with the current marine SQS biological endpoint, it is anticipated that the marine endpoint will be revised once the new Microtox® method is adopted.

The draft protocol calls for a control performance standard of at least 0.8 control final divided by control initial value (CF/CI), and a reference performance standard of at least 80% of the control. However, by inspection of the freshwater data set, it was found that a significantly larger percentage of the historic Microtox® data could be retained by modifying the control performance standard to 0.72 CF/CI, which was done for the purposes of this reliability assessment. This is considered reasonable, since the historic data were not collected using the new protocol. A more stringent performance standard may be applied for collection and evaluation of future data, or calculation of AETs.

3.3 Biological Hit/No-Hit List

Unlike AETs, the majority of the SQV sets being tested as part of the reliability assessment are not bioassay-specific; they are intended to identify levels associated with general policy goals, such as a "level below which adverse effects would not occur" or a "level above which adverse effects are likely to occur". Therefore, in this reliability assessment each station is assigned an overall biological hit/no-hit status that takes into account all the biological data available for that station. This will be referred to as a "pooled" endpoint. For this reason, the lowest AET and the lowest probable AET were selected as the pooled freshwater SQV sets that represent AETs, rather than individual Microtox® or *Hyalella* AETs.

To be consistent with the SQV sets being assessed, the overall pooled biological endpoint for a station should also be based on the combined biological test results for that station. Therefore, the biological hit/no-hit lists are based on the following definitions:

- **Hit:** At least one biological test shows an adverse response at that station
- **No-Hit:** No biological tests show an adverse response at that station

Most of the SQV sets were developed in pairs – a lower one representing a no adverse effect level, and a higher one representing levels above which effects are likely. The lower-level ones are similar in concept to the SQS, and the higher-level ones are similar in concept to the CSL. Therefore, two hit/no-hit definitions were developed for each set of SQVs - one based on the SQS definitions above, and one based on the CSL definitions above. In the case of the CSL definitions, both the one-hit and two-hit rules used for Washington State marine bioassays were applied; i.e., a CSL hit was assigned if any one bioassay endpoint exceeded its CSL level, or if two or more bioassay endpoints exceeded their SQS levels.

In addition, because of concerns over the relative lack of true chronic data in the database, a third hit/no-hit definition was developed to represent a statistical difference only, when compared to a control or

reference. If any biological test at a station showed an effect statistically greater than the control or reference sample, that station was deemed a "hit" station. Statistical significance is determined using standard SMS/PSDDA statistical protocols, at an alpha level of 0.05 for all bioassays. This hit/no-hit definition represents the most conservative end of the spectrum. Figure 3-1 shows the three levels of biological effects (statistical significance, SQS, and CSL) on a scale of increasing effects.

3.4 SEDQUAL Information System Modifications

To support the development of the hit/no-hit lists using the three definitions described above, two new analytical tools were developed for the SEDQUAL Information System. First, a tool was developed allowing the definition of custom freshwater (or marine) biological SQVs, called the Bioassay SQV Groups tool. This tool is found under the Utilities tab in the Group Definitions section, and allows the user to define groups of biological SQVs for a variety of tests and endpoints, with the user's choice of data preparation, statistical methods and transforms, hit definitions, and control and reference sample performance standards. Once one of these definitions has been developed, it can be used with the updated Bioassay Statistical Analysis (BSA) tool, the second SEDQUAL tool that was redeveloped to support this project.

The BSA tool can be found under the Analysis tab, and is similar to the previously-existing BSA tool, except that it includes a more user-friendly interface and additional features, and can be used with freshwater as well as marine data. The user selects the surveys, station groups, or sample groups for which the hit/no-hit analysis is to be completed, selects the biological tests and endpoints to be assessed, and selects the SQV definition to be used. Following this, the user may check the control or reference assignment for each sample, and then the hit/no-hit comparison is conducted.

The BSA tool was used to develop nine hit/no-hit lists – the three types of comparisons to reference and control for each of the three effects levels (statistical significance, SQS, and CSL) – which contained results for all of the bioassays and endpoints included in the runs. These results were downloaded into Excel files and the pooled hit/no-hit lists were developed by inspection, applying the pooling and one-hit/two-hit rules described above for each station. The final result of this analysis was a single station-by-station hit/no-hit list for each of the nine reliability runs.

3.5 Reliability Assessment

The project database was queried to obtain all chemistry data for the selected group of analytes, excluding any data qualified with a U, B, or X (see List of Acronyms for qualifier definitions). These data were downloaded into Excel workbooks, which are available from Ecology for review along with this report. There are nine Excel files, one for each combination of effects level and comparison method. The first worksheet, entitled "Notes," explains each of the worksheets in the file. The Worksheet "BioData" contains the hit/no-hit file downloaded from the SEDQUAL Information System, while the worksheet "BioHits" contains the final pooled hit/no-hit results as described above. The worksheet "ChemData" shows the chemistry data downloaded from the SEDQUAL Information System. A Visual Basic macro called **MakeTable** is then run to organize the data into a data table, as shown in the worksheet "DataTable." The "DataTable" worksheet also has a column into which the biological hit/no-hit values are entered for each station.

SAMPLE	Statistical Difference	SQS	CSL
A	Hit	No Hit	No Hit
B	Hit	Hit	Ho Hit
C	Hit	Hit	Hit

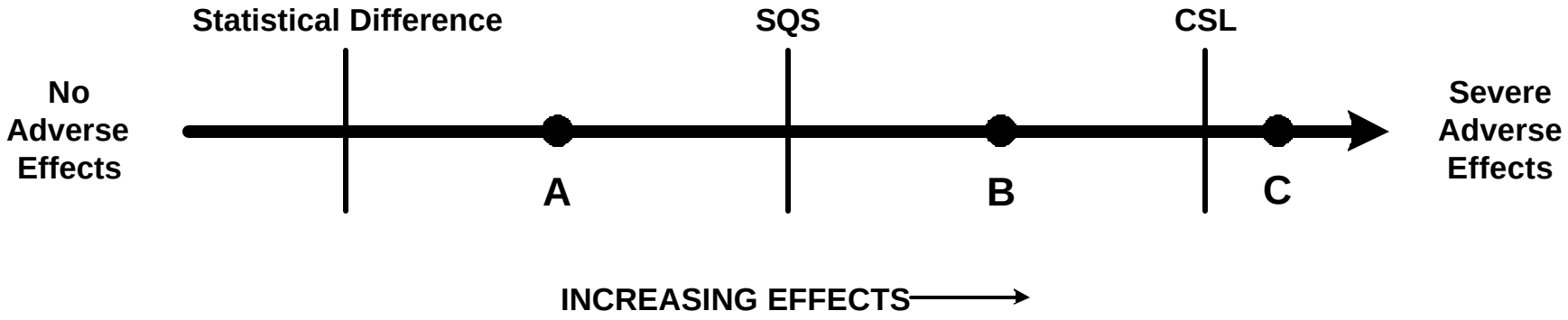


Figure 3-1. The Three Levels of Biological Effect

The next worksheet, entitled "Reliability," contains the criteria values for each of the eight SQV sets that were selected for evaluation, for the 46 analytes being assessed. These values are pre-entered in the right-hand part of the worksheet. To the left of these values, there are columns for false negatives, false positives, sensitivity, two definitions of efficiency, and overall reliability, which are calculated by a Visual Basic macro called **SQGReliability**. The **SQGReliability** macro compares the chemical concentrations of each chemical at a station to the corresponding criteria values in an SQV set, and determines whether a hit or no-hit would be predicted at that station. Then the chemical hit/no-hit prediction is compared to the biological hit/no-hit value, and the macro records whether the result is a correct prediction, a false positive, or a false negative. Once all stations and all chemicals have been assessed for a particular SQV set, the summary reliability statistics are entered into that row, and the macro moves on to the next SQV set. This and each of the other Excel macros were manually verified to ensure their accuracy.

The various reliability parameters are defined below and in the spreadsheets:

- **False Negatives:** hits predicted as no-hits/total number of hits
- **False Positives:** no-hits predicted as hits/total number of no-hits
- **Sensitivity:** hits correctly predicted/total number of hits (100% - % false negatives)
- **2002 Efficiency:** no-hits correctly predicted/total number of no-hits (100% - % false positives)
- **1988 Efficiency:** correctly predicted hits/total predicted hits
- **Reliability:** correct predictions/total stations

False positives and false negatives are the primary measure of predictive errors in the reliability assessment. Each of the other reliability values is related to them in some way. Most of these values can be compared across data sets and SQV types. However, because the denominator of the 1988 efficiency measure varies by SQV set and is not constant with respect to the data set, this measure cannot be compared across SQV sets, or against the results of 1997 freshwater AETs. It is used in this report mainly to look at the relative efficiency of different methods of comparison to control and reference, and to gain an overall sense of the accuracy of hit predictions.

The 1988 definition of efficiency is included because it has historically been used to assess AETs, and the 2002 measure was included because it is a more widely used measure that allows comparison across different SQV sets and is the mathematical counterpart to sensitivity. The difference between the two efficiency measures is subtle, since both are related to false positives, but they can be summarized as follows:

- 2002 Efficiency – how much of the clean area is identified as clean?
- 1988 Efficiency – how much of the area predicted to be toxic is actually toxic?

The main text of this report will focus on sensitivity, the two measures of efficiency, and overall reliability. False positives and false negatives, as well as all the backup information needed to calculate all the reliability values, can be found in the "Reliability" worksheets of the attached Excel workbooks, as well as in Appendix K.

3.6 Alternatives Assessment

The remaining worksheets in the Excel files were used in an exploratory manner to evaluate whether updating the freshwater AETs and/or calculation of alternative SQV sets would be likely to achieve a significant improvement in error rates. If so, a Phase II effort would be warranted. First, the macro **Distributions** was used to create hit and no-hit distributions for each chemical and enter them into the

"Distributions" worksheet. Next, the **CalcPercentiles** macro was used to calculate percentile values for the hit and no-hit distributions from one to 100, in increments of one, which were entered into the "Percentiles" worksheet. The **ErrorCalc** macro then calculated the six reliability measures for each percentile row, as if that percentile row were an SQV set.

Because the 100th percentile of the no-hit distribution is similar to the AETs, and the 95th percentile of the no-hit distribution is similar to the PAETs, this information can be used to determine whether updating the AETs with existing information would improve the error rates over the 1997 freshwater AETs and/or other SQV sets. In addition, other percentile values can be explored to identify the level that would provide the lowest error rates.

However, these are data sets with mixed bioassays, and AETs are normally calculated for individual endpoints. Therefore, one data set was selected (the SQS comparison to control), and AETs/PAETs were calculated for *Hyalella azteca* 10-day mortality, *Chironomus tentans* 10-day growth, *Chironomus tentans* 10-day mortality, and Microtox® luminescence (there were not enough data to calculate AETs for other endpoints). For this calculation, outliers were removed from the no-hit distributions by inspection using the 3x rule (if a concentration is three times higher than the next-lowest concentration, it is considered an outlier). Although a statistically-based outlier approach has been developed for use with the marine sediment data, attempts to apply this approach to the freshwater data set have met with less success because of larger gaps in the concentration distribution. The approach to outlier analysis can be revisited in Phase II, if the use of AETs is recommended; the other approaches to calculation of SQVs do not require outlier analysis.

The resulting AETs and PAETs were used to identify LAETs and PAETs for comparison to the 1997 values. It should be kept in mind that these were quick calculations done for screening purposes, and do not necessarily reflect the actual AETs that could be developed in Phase II. For this reason, the actual AET concentrations that were calculated are not provided in this report, although the results of the comparison are reported.

Finally, the **StartingCriteria** macro searched the percentile rows to find those values corresponding to various target false negative rates (5-30%, in increments of 5%), which also have the lowest possible false positive rates, and entered these percentile rows into the "FPCalc" worksheet. From there, the **FloatingPercentile** macro was used to adjust these values to optimize error rates and lower them even further. A complete discussion of the floating percentile method can be found in DEQ (1999), Appendix L, and Michelsen (1999), and additional materials can be requested from Avocet Consulting.

3.7 Results

In this section, the results of the database assembly, QA, and screening and the nine reliability runs are presented.

3.7.1 Data Set Descriptions

Two data sets were developed for the reliability runs, one containing all freshwater synoptic data remaining after screening and quality assurance, and one containing only those samples associated with valid reference stations. The number and types of bioassay endpoints in the final data sets are shown in Table 3-2, comprising 925 distinct sample/test combinations for the full data set, and 549 for the reference-only data set.

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

Test	No. of Samples Full Data Set	No. of Samples w/Valid Reference
Hyalella azteca 10-day mortality	383	216
Hyalella azteca 28-day mortality	7	0
Hyalella azteca 28-day growth	6	0
Chironomus tentans 10-day mortality	195	133
Chironomus tentans 10-day growth	162	130
Chironomus tentans 20-day mortality	16	16
Chironomus tentans 20-day growth	12	5
Microtox® decrease in luminescence	144	49

These samples are associated with 390 stations in the full data set and 239 stations with valid reference samples. Table 3-3 shows the number and percentage of stations associated with biological hits for each effects level and comparison method combination.

Table 3-3. Biological Hits for Each Reliability Run

Reliability Run	Biological Hits Number (Percent)
Statistical significance Comparison to control	235 (60%)
Statistical significance Comp. to mixed ref/control	227 (58%)
Statistical significance Comparison to reference	125 (52%)
SQS Comparison to control	182 (47%)
SQS Comp. to mixed ref/control	167 (43%)
SQS Comparison to reference	100 (42%)
CSL Comparison to control	113 (29%)
CSL Comp. to mixed ref/control	108 (28%)
CSL Comparison to reference	64 (27%)

From these results, it can be observed that comparison to the control is overall more conservative than comparison to reference, at each level of effect. However, it should be noted that this is not always the case for individual samples. In mortality tests, comparison to control is nearly always more conservative

than comparison to reference, because there is typically less mortality in the control sample than in the reference sample. However, in the acute growth tests represented in this data set, the organisms in the reference samples often grow larger than in the control samples (perhaps due to more food/organic matter in the reference samples). Therefore, a test sample may show a greater reduction in growth when compared to a reference sample than when compared to a control sample. The statistics above are likely influenced by the fact that the database has substantially more mortality and luminescence data than it has growth data.

3.8 Relative Reliability of Statistical Comparison Methods

One goal of doing the three different reliability runs for each effects level was to assess the comparative reliability of statistical comparisons to the control vs. comparisons to reference. The mixed case, where comparisons are made to the reference unless there is not a valid reference, in which case they are made to the control, was included because this represents an approach often taken to identifying hits in project data sets. If we must only include samples with valid reference stations, 40% of the data are lost, a substantial percentage. However, if these data are unreliable for use in developing SQVs, then it would be appropriate to screen them out.

Tables 3-3-a-l provide results for all reliability runs for these three methods of comparison, using the four primary reliability measures (sensitivity, 2002 efficiency, 1988 efficiency, and reliability). Results for the remaining reliability measures can be found in the backup Excel files, on the worksheets entitled "Reliability," which are also reproduced in Appendix K. Also listed on these sheets are the backup values needed to calculate the reliability measures, such as total number of stations, biological hits, biological no-hits, predicted hits, etc.

Examination of the above tables indicates that no one approach has greater reliability all the time; however, the mixed reference and control approach consistently had lower reliability than one of the other two approaches for all measures of reliability, all SQV sets, and all effects levels. On the other hand, in about 20% of the time it represented a compromise between the two, having intermediate reliability between control and reference.

Interestingly, sensitivity tended to be greater when comparing to reference, and both efficiency and reliability tended to be greater when comparing to the control. This is most likely because a comparison to reference generally reduces the number of hits in the data set, increasing the ability of most SQV sets to predict the hits that remain (i.e., increasing their sensitivity). It should be noted that just because sensitivity is greater in this context, does not necessarily mean that using a reference comparison is more conservative, since it is predicated on generating fewer biological hits in the data set.

Whether the reference or control has greater reliability also depends on which SQV sets are being examined, and which effects level is being addressed. As one would expect, comparison to control tends to be more reliable when used with SQV sets designed to be more conservative, and the comparison to reference tends to be more reliable when used with SQV sets designed to be less conservative. Likewise, comparison to control tends to be more reliable when applied to lower effects levels, and comparison to reference becomes increasingly more reliable when used with higher effects levels. Therefore, there is no one clear choice between these two approaches.

Table 3-3. Relative Reliability of Statistical Comparisons

a. Statistical Significance - Sensitivity

SQV Set	Control	Mixed	Reference
LAET	64	56	65
LPAET	68	60	68
TEL	96	93	92
PEL	59	54	69
TEC	89	83	86
PEC	52	44	55
LEL	95	88	83
SEL	47	41	52

b. Stat. Significance - 2002 Efficiency

SQV Set	Control	Mixed	Reference
LAET	62	50	51
LPAET	59	47	46
TEL	15	10	9
PEL	46	39	46
TEC	28	19	14
PEC	63	51	57
LEL	23	12	10
SEL	74	65	67

c. Stat. Significance - 1988 Efficiency

SQV Set	Control	Mixed	Reference
LAET	72	61	59
LPAET	71	61	58
TEL	63	59	53
PEL	62	55	58
TEC	65	59	52
PEC	68	56	58
LEL	65	58	50
SEL	74	62	63

d. Statistical Significance - Reliability

SQV Set	Control	Mixed	Reference
LAET	63	54	58
LPAET	64	54	57
TEL	64	58	52
PEL	54	48	58
TEC	64	56	52
PEC	56	47	56
LEL	67	56	48
SEL	58	51	59

e. SQS - Sensitivity

SQV Set	Control	Mixed	Reference
LAET	67	61	67
LPAET	69	63	70
TEL	96	92	92
PEL	60	60	71
TEC	87	81	88
PEC	53	50	57
LEL	95	87	84
SEL	50	48	55

f. SQS - 2002 Efficiency

SQV Set	Control	Mixed	Reference
LAET	58	52	50
LPAET	53	47	45
TEL	13	9	9
PEL	47	46	45
TEC	22	17	15
PEC	60	57	56
LEL	18	11	12
SEL	71	68	65

g. SQS - 1988 Efficiency

SQV Set	Control	Mixed	Reference
LAET	58	49	49
LPAET	56	47	48
TEL	49	43	42
PEL	50	46	48
TEC	49	43	43
PEC	54	46	48
LEL	50	42	41
SEL	60	53	53

h. SQS - Reliability

SQV Set	Control	Mixed	Reference
LAET	62	56	57
LPAET	60	54	55
TEL	51	44	44
PEL	53	52	56
TEC	52	45	46
PEC	57	54	56
LEL	54	44	42
SEL	61	59	61

i. CSL - Sensitivity

SQV Set	Control	Mixed	Reference
LAET	68	67	77
LPAET	70	69	80
TEL	96	94	95
PEL	70	70	84
TEC	88	86	94
PEC	62	62	73
LEL	95	91	91
SEL	58	58	67

j. CSL - 2002 Efficiency

SQV Set	Control	Mixed	Reference
LAET	52	51	50
LPAET	48	47	45
TEL	10	10	10
PEL	49	49	46
TEC	20	20	17
PEC	60	60	59
LEL	15	13	15
SEL	69	69	66

k. CSL - 1988 Efficiency

SQV Set	Control	Mixed	Reference
LAET	37	34	36
LPAET	35	33	35
TEL	30	28	28
PEL	36	34	36
TEC	31	29	29
PEC	39	37	40
LEL	31	29	28
SEL	44	42	42

l. CSL - Reliability

SQV Set	Control	Mixed	Reference
LAET	57	55	57
LPAET	54	53	54
TEL	35	33	33
PEL	55	55	56
TEC	40	38	37
PEC	61	60	63
LEL	38	35	35
SEL	66	66	66

3.9 Comparison Among SQV Sets

Tables 3-4a-i show the reliability results for all nine runs, for each of the eight SQV sets being assessed (as with Table 3-3, the backup data are provided in the "Reliability" worksheet of the associated Excel files, and in Appendix K).

The results show that the eight SQV sets consistently fall in the same order in terms of sensitivity, as follows (from most sensitive to least sensitive; range of sensitivities shown in parentheses):

- TEL (92-96%)
- LEL/TEC (81-95%)
- LPAET (60-80%)
- LAET/PEL (54-71%)
- PEC/SEL (41-73%)

AETs fall roughly in the middle of this group. The other SQV sets come in pairs, the lower of which is designed to be highly conservative and the upper of which is designed to indicate moderate to severe effects. Both the LPAET and the LAET are less sensitive than the lower values of all the other pairs and more sensitive than the higher values of all the other pairs.

2002 Efficiency, as would be expected, falls nearly in the opposite order, as follows (from most efficient to least efficient):

- SEL (65-74%)
- PEC/LAET (50-63%)
- LPAET (45-59%)
- PEL (39-49%)
- TEC (14-28%)
- LEL (10-23%)
- TEL (9-15%)

In this case, the AETs have moved from the middle up one step above the PELs. Percent efficiency ranges from around 10 for the TELs to the 70s for SELs. The trade-offs are not very good; high sensitivity is always accompanied by very low efficiency, and vice versa. However, the higher of each pair, along with both AET measures, do a better job of keeping both measurements above 50%.

1988 Efficiency tends to follow a similar pattern, but the differences between the various SQV sets are not nearly as significant. This measure of efficiency is much more strongly affected by the type of comparison (reference vs. control) and the hit/no-hit endpoint chosen (SQS vs. CSL) than by the SQV set used (see Table 3-4). More conservative comparisons and more conservative measures of effects result in higher 1988 efficiency, because a hit predicted by chemical criteria is more likely to turn out to be an actual hit when the biological data are assessed in a more conservative manner. 1988 Efficiency ranges from 62-74% in the statistical comparison to control run, down to 28-40% in the CSL comparison to reference.

Overall reliability results are more variable, with no one SQV set ranking highest most of the time. The group of SQV sets that has higher reliability varies with the level of effects that is being assessed. However, AETs are always in the higher-scoring group, showing that they present a relatively good trade-off between sensitivity and efficiency. For statistical significance only, the higher scoring group consists

Table 3-4. Results of Reliability Runs for Eight SQV Sets

a. Statistical Significance - Control

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	96	15	63	64
LEL	95	23	65	67
TEC	89	28	65	64
LPAET	68	59	71	64
LAET	64	62	72	63
PEL	59	46	62	54
PEC	52	63	68	56
SEL	47	74	74	58

b. Statistical Significance - Mixed

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	93	10	59	58
LEL	88	12	58	56
TEC	83	19	59	56
LPAET	60	47	61	54
LAET	56	50	61	54
PEL	54	39	55	48
PEC	44	51	56	47
SEL	41	65	62	51

c. Statistical Significance - Reference

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	92	9	53	52
LEL	83	10	50	48
TEC	86	14	52	52
LPAET	68	46	58	57
LAET	65	51	59	58
PEL	69	46	58	58
PEC	55	57	58	56
SEL	52	67	63	59

d. SQS - Control

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	96	13	49	51
LEL	95	18	50	54
TEC	87	22	49	52
LPAET	69	53	56	60
LAET	67	58	58	62
PEL	60	47	50	53
PEC	53	60	54	57
SEL	50	71	60	61

e. SQS - Mixed

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	92	9	43	44
LEL	87	11	42	44
TEC	81	17	43	45
LPAET	63	47	47	54
LAET	61	52	49	56
PEL	60	46	46	52
PEC	50	57	46	54
SEL	48	68	53	59

f. SQS - Reference

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	92	9	42	44
LEL	84	12	41	42
TEC	88	15	43	46
LPAET	70	45	48	55
LAET	67	50	49	57
PEL	71	45	48	56
PEC	57	56	48	56
SEL	55	65	53	61

g. CSL - Control

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	96	10	30	35
LEL	95	15	31	38
TEC	88	20	31	40
LPAET	70	48	35	54
LAET	68	52	37	57
PEL	70	49	36	55
PEC	62	60	39	61
SEL	58	69	44	66

h. CSL - Mixed

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	94	10	28	33
LEL	91	13	29	35
TEC	86	20	29	38
LPAET	69	47	33	53
LAET	67	51	34	55
PEL	70	49	34	55
PEC	62	60	37	60
SEL	58	69	42	66

i. CSL - Reference

SQV Set	Sensitivity	2002 Eff.	1988 Eff.	Reliability
TEL	95	10	28	33
LEL	91	15	28	35
TEC	94	17	29	37
LPAET	80	45	35	54
LAET	77	50	36	57
PEL	84	46	36	56
PEC	73	59	40	63
SEL	67	66	42	66

of the lower of all the pairs (TEL, TEC, LEL) and the two AET measures. However, the differences are not great and largely disappear when comparison is made to reference only. For the SQS and CSL levels, the higher scoring group includes the higher of all the pairs (PEL, PEC, SEL) and the two AET measures. Percent reliability is in general not very high for any of the measures, ranging from the mid-30s to the mid-60s.

3.10 Recalculation of AETs

A preliminary recalculation of one set of AETs was conducted as an example, to determine how the values might change and whether or not they would exhibit increased reliability. The SQS level of effects was selected, since it would have an intermediate number of hits in the data set and was designed to be similar to a level used in Ecology's regulatory programs (it may or may not actually be similar, due to the relative lack of chronic data in this data set compared to the marine data set). The comparison to control was selected because this comparison uses the larger data set and is more reliable than the mixed reference and control comparisons. AETs and PAETs were calculated for four endpoints described in the Methods section: *Hyalella azteca* 10-day mortality, *Chironomus tentans* 10-day mortality, *Chironomus tentans* 10-day growth, and Microtox® luminescence. These 2002 LAETs and PAETs were compared to the 1997 LAETs and PAETs in terms of the six reliability measures, using a copy of the same spreadsheet used to evaluate the reliability of the original eight SQV sets for SQS comparisons to control.

The results of this comparison are shown in Table 3-5, below. 2002 LAET and LPAET values generally have lower sensitivity, higher efficiency, and slightly lower overall reliability than the 1997 values, even though two new chironomid endpoints were added which would be expected to be relatively conservative. Most likely, the decrease in sensitivity was due to the higher Microtox® values in the 2002 AETs, which increased because of the additional of new, higher no-hit data and the screening out of some older data that did not pass quality assurance. Table 3-5 also shows the results when the 2002 *Hyalella* and *Chironomus* AETs/PAETs are combined with the 1997 Microtox AETs/PAETs. This results in a more sensitive SQV set, but there is a corresponding loss of efficiency.

Table 3-5. Reliability of 1997 and 2002 AETs/PAETs for SQS Comparison to Control

SQV Set	Sensitivity	Efficiency	Reliability
1997 LAET	67	58	62
1997 LPAET	69	53	60
2002 LAET	52	63	57
2002 LPAET	65	52	58
Mixed LAET	72	48	59
Mixed LPAET	75	43	58

Overall, recalculation of AETs/PAETs at this time does not appear to increase reliability over use of the 1997 values. It should be kept in mind that this is based on one limited example, and further evaluation may change this conclusion. It is also likely the case that the freshwater data set is inherently more variable than the marine data set, due to the much wider variety of geochemical and biological environments within Washington State. Stratification of the data set to reflect these variations may produce better results.

Other percentiles of the no-hit distribution could potentially be explored as well. Perusal of the Percentile worksheets for all nine reliability runs indicates that the optimal percentile (all reliability measures above 60% and a good balance between sensitivity and efficiency) is always in the no-hit distribution, typically between the 85th and 93rd percentile of the full no-hit distribution (no outliers removed). Percentiles of the

statistical significance only with comparison to control have by far the best reliability rates, over 70% for all three measures of reliability. This might be the most appropriate hit definition and distribution to work with, because of its relatively low error rates and more conservative nature, potentially making up for the lack of chronic tests in the current suite of biological tests. Reliability is most likely highest for this particular run because the direct comparison to control reduces variability among data sets.

3.11 Floating Percentile Calculations

Table 3-6 shows the best results obtained by the Floating Percentile method for each of the nine reliability runs. More complete results can be found in the attached Excel files, in the "FPCalc" Worksheets, each of which are also reproduced in Appendix L. This method requires you to choose among various sensitivity ranges, and then optimizes the efficiency for each sensitivity level. The target sensitivity for these results was at least 70%, with efficiency and reliability levels of at least 60%. In most cases this could be achieved, and in some cases it was possible to increase all three measures above 70%. The "reduction in errors" column adds together reductions in false positive rates and false negative rates, while the "improvement in reliability" column indicates the improvement in overall reliability observed. The two values are not the same because there is not an additive relationship between the three measures; they do not all have the same denominator (see Methods section for definitions).

Table 3-6. Results of Floating Percentile Calculations

Reliability Run	Sensitivity	Efficiency	Reliability	Reduction in Errors	Improvement in Reliability
STAT control	75	70	73	-8	+3
STAT mixed	70	41	58	-10	+4
STAT reference	70	61	66	-13	+6
SQS control	70	63	67	-9	+5
SQS mixed	70	52	60	-8	+5
SQS reference	70	58	63	-12	+7
CSL control	71	73	73	-13	+11
CSL mixed	70	66	67	-10	+8
CSL reference	75	69	70	-15	+10

Regardless of the level of effects being evaluated, the best results using this method are with comparison to the control, although comparison to reference also works well at the CSL level. Comparison to the mixed reference and control is always the worst-performing choice.

The Floating Percentile method also identifies chemicals that may not be contributing to toxicity in the data set. Those chemicals that can be increased to their maximum value without affecting false negative rates are unlikely to be contributing to toxicity within their existing concentration distributions. The reliability of any of the SQV sets, whether existing, newly calculated AETs, or created by an alternative method, will likely be improved by removing these chemicals from the assessment. For this regional data set, the chemicals fall out as follows:

- **Primary Importance:** Antimony, cadmium, chromium, copper, mercury, nickel, silver, zinc, TOC, bis(2-ethylhexyl)phthalate, di-n-butyl phthalate, PCBs, PAHs (summed measures).
- **Secondary Importance:** Lead, ammonia, sulfides, individual PAHs, Dieldrin, Endrin, hexachlorobenzene, phenol, DDTs.

- **Not Contributing:** 2,3,7,8-TCDD, Aldrin, gamma-BHC.

Patterns within the results strongly suggest that a combined measure of PAHs would be preferable to SQVs for individual PAHs, as these individual and summed measures are strongly correlated and may be acting in an additive or synergistic fashion.

3.12 Conclusions and Recommendations

As Phase I of an assessment of freshwater SQV sets for use in Ecology's sediment management programs, a regional database of synoptic freshwater chemistry and bioassay data was assembled, and the SEDQUAL Information System was updated to incorporate tools to allow development of user-defined biological SQVs and to conduct statistical hit/no-hit analyses of freshwater bioassay data. Existing freshwater SQV sets in North America were screened, and the best of these SQV sets were selected for reliability assessment against the regional freshwater database. These SQV sets included the 1994 AETs and PAETs, the TELs and PELs, Ontario's LELs and SELs, and the consensus-based TECs and PECs. Each of the eight SQV sets was evaluated for six reliability parameters (the three main ones are reported here), at three levels of effects and using three different methods of comparison to control and/or reference. Finally, an exploratory assessment of recalculation of AETs and/or other methods of calculating freshwater values was conducted.

3.12.1 PHASE I CONCLUSIONS

The detailed results of this reliability assessment are contained in a set of associated spreadsheets, and the primary conclusions are as follows:

- The eight SQV sets followed a consistent pattern of conservativism, with the TELs, LELs, and TECs being the most sensitive, the AETs and LAETs falling in the middle, and the PELs, SELs, and PECs being the least sensitive. Efficiency results were generally the opposite of these trends.

- The most conservative SQV sets (TELs/TECs/LELs) have very high rates of false positives, typically 80-90%, and would therefore not be very useful even for screening out areas that do not need to be assessed, as stations would almost never be screened out.

- The PECs and SELs conversely had low sensitivity, typically below 60% (false negative rates of 40% or higher). These levels are likely not conservative enough to use as screening levels or cleanup standards.

- The AETs/PAETs and the PELs had the best balance of sensitivity and efficiency, with typically lower error rates than the other SQV sets. However, even these SQV sets had sensitivity typically ranging from 60-80%, and efficiency from 40-60%. Both values are somewhat low for use in a regulatory setting. Any of these SQV sets would almost always require confirmatory bioassays.

- Comparison to control appears to be the method capable of producing the most reliable SQV sets, followed by comparison to reference. The use of mixed reference and control comparisons produces low reliability and is not recommended, either for SQV development or for interpretation of data.

- Recalculation of the AETs using the 2002 data set does not appear to improve reliability rates. Both the sensitivity and overall reliability decrease slightly, most likely based on the addition of higher no-hit data for the Microtox® bioassay, even though two additional Chironomid endpoints have been

added. However, this should be considered a tentative conclusion, as it is based only on a single hit/no-hit definition (SQS) with comparison to control, and relatively simple outlier procedures.

3.12.2 PHASE II RECOMMENDATIONS FOR FUTURE WORK

- Certain chemicals are apparently not contributing to toxicity in the data set, and reliability may be improved by removing them. A more significant improvement may be achieved by summing the individual PAHs into one or two aggregate measures, better reflecting their mode of toxicity and behavior in the data set.

- Organic-carbon normalization could be explored to see whether it increases the reliability of the either recalculated AETs or floating percentile values, although it has not had a significant effect in the past.

- It would be worthwhile to further explore the recalculation of AETs, at least to confirm the preliminary results noted above, using any improvements obtained from the first two bullets. If this appears to be a viable approach, calculate individual AETs and assess their reliability separately, as well as on a pooled basis. If Phase I results are verified (i.e., reliability is not improved significantly), evaluate one or both of the options below.

- Selection of a somewhat lower percentile of the no-hit distribution (85^{th}-93^{rd}) might provide better reliability results, slightly weighted toward greater sensitivity. In addition, use of the Floating Percentile method of SQV optimization along with the optimal percentiles should be considered, as it appears capable of producing criteria sets with reliability of greater than 70% for all measures.

- Use of an exceedance ratio approach could be evaluated, to assess the overall toxicity of a station rather than chemical-by-chemical toxicity.

Although it would be desirable to look at acute and chronic effects separately, and to calculate chronic AETs or other SQV sets, there is not yet enough chronic data in the database with which to conduct these analyses. Similarly, there is not yet adequate benthic data (synoptic with chemistry data) to calculate benthic SQVs or conduct reliability assessments of other SQVs against benthic community data. In the coming years, it is recommended that Ecology emphasize the collection of synoptic freshwater chemistry, chronic bioassay, and benthic data to ensure that chronic effects are represented in the data set. The collection of more benthic community data can best be facilitated by developing accepted, standardized protocols for the collection and interpretation of this type of data. In the near term, the following additional activities are recommended for Phase II to support collection of benthic data:

- A workshop to address sample collection, taxonomic identification, and other protocol issues specific to freshwater sediments.

- Development of endpoints for assessment of freshwater benthic community data, including whether there is an appropriate role for rapid assessment techniques.

3.12.3 RECOMMENDATIONS FOR USE IN REGULATORY PROGRAMS

None of the existing SQV sets is ideal for direct use in regulatory programs, as each has limited reliability when compared with regional freshwater data. The Phase II recommendations above are designed to produce SQV sets with improved reliability that could potentially be used as stand-alone screening levels

or cleanup levels, as part of a larger program that also includes biological testing alternatives. In the meantime, the following recommendations are offered for using existing SQV sets:

- **Lower Screening Level –** TELs, TECs, and LELs are often used (along with ERLs developed by NOAA) in North America as screening levels for determining whether sediments require further attention. The reliability results obtained above suggest that these values can be used in the Pacific Northwest to screen areas out from further regulatory attention, but not to screen them in. If sediments do not exceed these values, they are unlikely to exhibit any adverse effects in bioassay tests. Unless such sediments have potential bioaccumulation issues, they can be considered clean, requiring no further action and suitable for open-water disposal, habitat restoration, and other beneficial uses.

  However, it should be noted that these values are not effective for screening areas in – that is, for identifying areas that require further attention. Because they have very low efficiency (10-20%), 80-90% of the time areas screened in by these values will turn out to have no adverse effects and require no cleanup. None of the existing SQV sets were found to be effective for this purpose, and this is recommended as a focus of attention in Phase II.

- **Upper Screening Level –** In most programs, it is helpful to have levels above which it is nearly certain that adverse effects will be observed. Among the SQV sets evaluated, the SEL was the level above which adverse effects are most likely, but there is still a 25-50% chance (depending on the hit definition) that there will not be adverse effects above this level. Therefore, this level should not be used as a hard and fast requirement for cleanup or prohibition from open water disposal, but perhaps a strong indication that effects are likely, and a requirement for biological testing if the regulated party wishes to rebut that presumption. An upper screening level is also recommended for development in Phase II.

- **Prioritizing Areas for Attention –** Currently, there is a large gray area between the above two levels, and a need to make decisions in many freshwater areas. Although the SQVs are not yet reliable enough to use on a stand-alone basis, they could be used to help prioritize areas for attention. For example, areas exceeding the SEL could be considered the highest-priority areas for biological testing. Areas between the PAET and the SEL could be considered medium priority for attention, and areas between the TEL and the PAET could be considered low priority. Areas below the TEL would be screened out entirely.

4.0 REFERENCES

Detailed references for the SQV methods can be found in their respective technical appendices.

Andreasson, J. 1991. Weyerhaeuser-Longview Pulp and Paper Mill Class II inspection. Washington State Department of Ecology, Olympia, WA.

ASTM. 2000. Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. ASTM E1706-00. American Society for Testing and Materials, West Conshohocken, PA.

Bennett, J. and J. Cubbage. 1992a. Copper in sediments from Steilacoom Lake, Peirce County, Washington. Washington State Department of Ecology, Olympia, WA.

Bennett, J. and J. Cubbage. 1992b. Effects of polycyclic aromatioc hydrocarbons in sediments from Lake Washington on freshwater bioassay organisms and benthic macroinvertebrates. Washington State Department of Ecology, Olympia, WA.

BNSF. 2002. BNSF Skykomish River Site. July 2001. Citation obtained from SEDQUAL , Survey ID: BNSFSK02.

Century West Engineering Corporation. 1989. Report of findings: Phase II Columbia River Impact Investigation, Vancouver, WA. Prepared for the Port of Vancouver, Vancouver, WA, by Century West Engineerning Corporation.

CRD. 2000. The Trigger Process: An Early Warning of Adverse Environmental Effects for the Clover and Macaulay Point Outfalls. Appendix to the Liquid Waste Management Plan, Stage 3. Capital Regional District, Victoria, BC.

Cubbage, J. 1992. Survey of contaminants in sediments in Lake Union and adjoining waters. Washington State Department of Ecology, Olympia, WA.

Cubbage, J., D. Batts, and S. Breidenbach. 1997. Creation and Analysis of Freshwater Sediment Quality Values in Washington State. Prepared for the Sediment Management Unit, Washington Department of Ecology, Olympia, WA.

Dames and Moore. 1991. Columbia Slough sediment analysis and remediation project. Prepared for City of Portland by Dames and Moore, Portland, OR.

Dames and Moore. 1997. Portland Shipyard Environmental Audit, Cascade General. Portland Shipyard Environmental Audit Appendix D Dames & Moore Data. Citation obtained from SEDQUAL , Survey ID: PSYD&M97.

Dames and Moore. 1998. Willamette River Data. Portland Shipyard Environmental Audit Appendix D Dames & Moore Data. Citation obtained from SEDQUAL , Survey ID: WRD&M98.

Das, T. 1991. Longview Fibre Company Class II inspection. Washington State Department of Ecology, Olympia, WA.

Degens. 1998. Sediment Characterization Study, Marine Terminal 2 Berths 203-206 and Marine Terminal 4 Berth 416. Citation obtained from SEDQUAL , Survey ID: PPTLDT24.

DEQ. 1999. Portland Harbor Sediment Management Plan, Appendix G. Oregon Department of Environmental Quality, Portland, OR.

Ecology. 1988. Screening Survey For Chemical Contaminants and Toxicity in Sediments at Five Lower Columbia River Ports. Prepared by Washington State Department of Ecology, Olympia, WA.

Ecology. 1990. Port of Vancouver results of Daphnia magna sediment bioassays. Prepared by Washington State Department of Ecology, Olympia, WA.

Ecology. 1994a. Everett Simpson Site Sediment Investigation. June 1994. Citation obtained from SEDQUAL , Survey ID: EVRTSM94.

Ecology. 1994b. Spokane River PCB Bioassay Study. August 1994. Citation obtained from SEDQUAL , Survey ID: SPOKNR94.

Ecology. 1995. Sediment Sampling and Analysis Plan Appendix: Guidance on the Development of Sediment Sampling and Analysis Plans Meeting the Requirements of the Sediment Management Standards—Chapter 173-204 WAC. Prepared for the Washington Department of Ecology, Olympia, WA, by PTI Environmental Services, Bellevue, WA.

Ecology. 1997. Tri-Star Marine NPDES Sediment Monitoring. Citation obtained from SEDQUAL , Survey ID: TRI-STAR.

Ecology. 2000. Review of Sediment Quality Data for the Similkameen River. Washington State Department of Ecology, Olympia, WA. July 2000.

Ecology. 2001. Lake Roosevelt Sediment TMDL Study. Reassessment of Toxicity of Lake Roosevelt Sediments. November 2001, Draft. Washington State Department of Ecology, Olympia, WA. July 2000.

Ecology. 2002. Microtox® 100 Percent Sediment Porewater Toxicity Assessment. Draft Final. Washington Department of Ecology, Olympia, WA.

Ecology and Environment. 2001. McCormick and Baxter Creosoting Company sediment remedial design, Phases I and II. Prepared for Oregon Department of Environmental Quality, Portland, OR, by Ecology and Environment, Portland, OR.

Embry, S. S. and E. K. Block. 1995. Reconnaissance investigation of water quality bottom sediment, and biota associated with irrigation drainage in the Columbia Basin Project, Washington, 1991-92. U.S. Geological Survey, Tacoma, WA.

ENSR. 1994a. Aluminum Company of America. Vancouver Works. Baseline Sediment Characterization. ENSR document 0225-006-405.

ENSR. 1994b. Columbia Aluminum Company. Goldendale, Washington. Baseline Sediment Characterization. ENSR document 1774-004-500.

Exponent. 2001. Sediment Sampling Results for Quendall Terminals Property. Prepared for the City of Renton, WA, by Exponent, Inc., Bellevue, WA.

FishPro and GeoEngineers. 1991. Biological report on sediment and water bioassays and benthic community determination at Unimar Yard 1 dry dock facility, Seattle, WA. Prepared for Unimar International, Inc, Seattle, WA, by FishPro, Port Orchard, WA, and GeoEngineers, Redmond, WA.

Fore, L.S. 2001. Evaluation of Alternative Sampling Designs for Biological Monitoring of Streams. Prepared for the Washington Department of Ecology, Olympia, WA.

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

Friedman and Bruya, Inc. 1990. Marco Shipyard sediment monitoring results. Memo from Friedman and Bruya, environmental chemists to Doc Church, MARCO Shipyards, Seattle, WA.

Golder Associates. 1992. Sediment sampling and analysis report for Cedar River Delta sediments. Prepared for City of Renton, WA, by Golder Associates, Redmond, WA.

Harding ESE. 2001. Cargill Irving Elevator Terminal; Dredged material characterization for disposal at Ross Island, Portland, OR. Prepared for Oregon Department of Environmental Quality, Portland, OR, by Harding ESE, Novato, CA.

Hart Crowser. 1992. Lake Union Drydock, sediment monitoring program results. Prepared fro Lake Union Drydock Company, by Hart Crowser, Inc., Seattle, WA.

Hart Crowser. 2000. Port of Portland Ross Island Site Investigation. Prepared for the Port of Portland by Hart Crowser, Inc., Seattle, WA.

Hart Crowser. 2002. Lower Willamette River Reference Area Study. Prepared for the U.S. Army Corps of Engineers, Portland District, Portland, OR, by Hart Crowser, Inc., Lake Oswego, OR.

Heffner, M. 1989. Kalama Chemical, Inc. Class II inspection. Washington State Department of Ecology, Olympia, WA.

Heffner, M. 1991. Reynolds Metal Company Class II inspection. Washington State Department of Ecology, Olympia, WA.

Johnson, A. 1991. Review of metals, bioassay, and macroinvertebrate data from Lake Roosevelt benthic samples collected in 1989. Memo to Carl Nuechterlein. Washington Department of Ecology, Olympia, WA.

Johnson, A. and D. Norton. 1988. Screening survey for chemical contaminants and toxicity in drainage basins at Paine Field. Washington State Department of Ecology, Olympia, WA.

Johnson, A. and D. Norton. 2001. Chemical analysis & toxicity testing of Spokane River sediments. Prepared for Washington State Department of Ecology, Lacy, WA.

Johnson, A., D. Norton, and B. Yake. 1989. An assessment of metals contamination in Lake Roosevelt. Washington State Department of Ecology, Olympia, WA.
Johnson, A., and M.Heffner. 1993. Boise Cascade mill at Walulla, Class II inspection. Washington State Department of Ecology, Olympia, WA.

Kagley, A.N, K.E. Kardong, R.G. Snider, and E. Casillas. 1995. Effects of chemical contaminants on growth, age structure, and reproduction of *Mytilus edulis* complex from Puget Sound, Washington. *Proceedings of Puget Sound Research '95*:678-684.

King County. 1999. Lake Sammamish Baseline Sediment Study. August 1999. Citation obtained from SEDQUAL , Survey ID: LSAMM99.
King County. 2000a. Lake Washington Baseline Sediment Study 2000. August 2000. Citation obtained from SEDQUAL , Survey ID: LKWA00.

King County. 2000B. Lake Union University Regulator CSO Post Separation Study 2000. April 2000. Citation obtained from SEDQUAL , Survey ID: LUUCSO00.

Landau Associates. 1993. Mill Creek and East Drain Sediment Sampling Report, Western Processing, Phase II, Kent, WA. Prepared for Western Processing by Landau Associates, Edmonds, WA.

Michelsen, TC. 1999. Error Rate Minimization Techniques for Calculating Sediment Quality Guidelines. Interactive Poster Presentation at SETAC '99, Philadelphia, PA.

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

Norton, D. 1991. Distribution and significance of Polycyclic Aromatic Hydrocarbons in Lake Washington sediments adjacent to Quendall Terminals/J.H. Baxter Station. Washington State Department of Ecology, Olympia, WA.

Norton, D. 1992. Results of sediment sampling in Baxter Cove, Lake Washington. Washington State Department of Ecology, Olympia, WA.

Parametrix. 1993. Results of acute toxicity tests on freshwater sediments collected from Silver Lake, WA, using *Hyalella azteca* and Microtox. Prepared for Kramer, Chin & Mayo, Inc., by Parametrix, Inc., Kirkland, WA.

PTI. 1989. Puget Sound Dredged Disposal Analysis Guidance Manual: Data Quality Evaluation for Proposed Dredged Material Disposal Projects. Prepared for Washington Department of Ecology, Olympia, WA, by PTI, Environmental Services, Bellevue, WA.

PTI. 1992. McCormick and Baxter Creosoting Company. Remedial Investigation Report, Phase 1 and Phase 2. Prepared for the Oregon Department of Environmental Quality, Portland, OR.

Quinn. 1998. Terminal 4 Slip 3 Sediment Investigation. Marine Terminal 2 Berths 203-206 and Marine Terminal 4 Berth 416. Citation obtained from SEDQUAL , Survey ID: WLRPT498.

Reif, D. 1989. Pacific Woodtreating Corporation, Ridgeway, WA, Class II inspection. Washington State Department of Ecology, Olympia, WA.

Reynoldson, T.B. and K.E. Day. Undated. Biological Guidelines for the Assessment of Sediment Quality in the Laurentian Great Lakes. Environment Canada, National Water Research Institute, Burlington, ON.

Reynoldson, T.B., R.H. Norris, V.H. Resh, K.E. Day, D.M. Rosenberg. 1997. The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16(4):833-85.

Ruiz, C. 1989. Ferndale Wastewater Treatment Plant Class II Inspection. Washington State Department of Ecology, Olympia, WA.

SAIC. 1991. Station Hazard Assessment Report; Hansville Landfill Kitsap County, Washington Source: Elaine Atkinson, Washington State Department of Ecology NWRO-TCP, Bellevue, WA.

SEA. 1998. Portland Shipyard Sediment Investigation Data Report. Prepared by Striplin Environmental Associates, Olympia, WA. Citation obtained from SEDQUAL , Survey ID: PSYSEA98.

Serdar, D., J. Cubbage, and D. Rogowski. 2000. Concentrations of chemical contaminants mand bioassay response to sediments in Salmon Bay, Seattle; results of phase III sampling. Prepared by Washington Department of Ecology, Olympia, WA.

Shannon and Wilson. 1994. Sediment sampling report Seattle Commons Parcel C. Prepared for Seattle Commons by Shannon and Wilson, 400 N 34th ST, Seattle, WA.

Tetra Tech. 1993. Lower Columbia Backwater Recon. Survey. June, 1993. Citation obtained from SEDQUAL , Survey ID: LCBWRS93.

TOSCO. 1997. TOSCO Sediment Sampling Results 1999. Citation obtained from SEDQUAL , Survey ID: TOSCO99.

USGS. 1993. Determination of Miscellaneous Metals…from Milltown Reservoir and Clark Fork River, Montana. 1991. U.S. Geological Survey, Montana.

USGS. 1992. Sediment-Quality Assessment of Franklin D. Roosevelt Lake and the Upstream Reach of the Columbia River, Washington. September 1992, revised 1994.

Widdows, J., and Donkin, P. 1992. Mussels and Environmental Contaminants: Bioaccumulation and Physiological Aspects. In: The Mussel *Mytilus*: Ecology, Physiology, Genetics and Culture, Developments in Aquaculture and Fisheries Science. E. Gosling, Ed., Elsevier, Amsterdam, pp. 383-424.

Yake, B., D. Norton, and M. Stinson. 1986. Application of the triad approach to freshwater sediment assessment: an initial investigation of sediment quality near Gas Works Park, Lake Union. Washington State Department of Ecology, Olympia, WA.

Zinner, L. 1990. Aluminum Company of America Class II inspection. Washington State Department of Ecology, Olympia, WA.