Sediment Quality Values Refinement: Volume I

1988 UPDATE AND EVALUATION OF PUGET SOUND AET

FINAL REPORT

September 1988

Prepared for
U.S. Environmental Protection Agency
Region 10 - Office of Puget Sound
Seattle, Washington

PTI Environmental Services
3625 132nd Avenue SE
Suite 301
Bellevue, WA 98006

VOLUME I

SEDIMENT QUALITY VALUES REFINEMENT: 1988 UPDATE AND EVALUATION OF PUGET SOUND AET

by

R. Barrick, S. Becker, L. Brown, H. Beller, and R. Pastorok

Submitted to

Tetra Tech, Inc.
11820 Northup Way, Suite 100
Bellevue, WA 98005

For

Puget Sound Estuary Program
Office of Puget Sound, Region 10
U.S. Environmental Protection Agency

Funded by

National Estuary Program
Office of Marine and Estuarine Protection
U.S. Environmental Protection Agency

EPA Contract No. 68-01-4341
PTI Contract No. C717-01

September 1988

EXECUTIVE SUMMARY

During 1986 and 1987, the Office of Puget Sound, U.S. EPA Region 10, the U.S. Army Corps of Engineers Seattle District, and the Washington Departments of Ecology and Natural Resources undertook a joint effort to develop sediment quality values for Puget Sound. The goal of the project was to identify concentrations of chemicals in sediments that are expected (based on field evidence or theoretical predictions) to be associated with adverse biological effects. The specific objectives of the study were to compile synoptic biological and chemical data from Puget Sound, to evaluate techniques that could be used to develop chemical-specific sediment quality values, and to evaluate the reliability (i.e., ability to correctly identify sites with known biological effects) of the values generated using different techniques. The study was completed in September 1986 and copies of the final report (Beller et al. 1986) were distributed nationally and in the Puget Sound region.

The 1986 report contains an evaluation of several different approaches to establishing sediment quality values and ranges of chemical values that can be applied in sediment management, but does not identify the specific methods or values that the agencies will adopt for regulation or how these numerical values should be modified for application in specific regulatory programs.

Based on the conclusions of the 1986 report, the agencies requested a more detailed evaluation of the use of the Apparent Effects Threshold (AET) approach in establishing sediment quality values. An AET is defined as the sediment concentration of a contaminant above which statistically signficant ($P \leq 0.05$) adverse effects for a particular biological indicator are always expected relative to appropriate reference conditions. AET are considered distinct from sediment quality values. AET are tools that can be used in sediment management. Sediment quality values, which may be based in part on AET, have yet to be adopted by Puget Sound agencies and, as a policy decision, may include safety factors or other modifications. This report does not address policy decisions for use of AET by agencies. It is recommended that the AET values not be applied outside Puget Sound without further evaluations.

The overall objective of this report is to further test the reliability of AET and potential resulting sediment quality values based on AET in predicting adverse biological effects, and to provide more detailed analyses of the relationships between chemical concentrations and biological effects. Most of these evaluations were conducted as part of the 1988 recalcuation and evaluation of Puget Sound AET presented in this report (Volume I). Other analyses conducted for the Puget Sound Dredged Disposal Analysis program are summarized in a separate volume (Barrick et al. 1988; Volume II).

The following six objectives were accomplished as tests of the reliability of AET:

1. Further test the predictive reliability of 1986 Puget Sound AET using new data sets for Eagle Harbor, Elliott Bay, and Everett Harbor

2. Evaluate the predictive reliability of Puget Sound AET after incorporation of these new data sets in the calculation of 1988 AET

3. Compare the predictive reliability of 1986 and 1988 Puget Sound AET using independent dredging data sets (only a limited comparison was possible because of incomplete chemical or biological test results for most of these dredging studies)

4. Investigate the characteristics of and develop hypotheses for stations at which observed biological effects run counter to AET predictions

5. Test the influence of geographic location on AET reliability

6. Determine if a subset of chemicals in the sediment quality database accounts for all or nearly all of the predictive reliability of AET.

The following three additional objectives were accomplished by applying the sediment quality values database to Puget Sound data:

1. Investigate minimum data requirements to derive reliable AET

2. Identify potential problem areas by predicting biological effects in Puget Sound from chemical data

3. Evaluate the influence of different bioassay reference data sets on the development of AET.

Evaluations conducted for this PSEP study indicate that 1988 AET values are generally reliable predictors of adverse biological effects after recalculating the 1986 AET values by incorporating new data from Eagle Harbor, Elliott Bay, and Everett Harbor. Of the 201 benthic infauna stations and 287 amphipod bioassay stations evaluated for 13 Puget Sound embayments, approximately 85 percent (174 stations and 243 stations, respectively) are in accordance with the predictions of the 1988 AET values for these indicators (i.e., they do not exhibit adverse effects at chemical concentrations less than the AET values, and do exhibit adverse effects at chemical concentrations above the AET values). One or the other of these biological indicators is still in accordance with predictions of the respective AET values at 23 of the 28 (82 percent) stations at which an impact was not predicted and data for both benthic infauna and amphipod bioassays were available. Because the 1988 AET are based on data for all available nonimpacted stations, none of the 91 nonimpacted benthic stations or 177 nonimpacted amphipod bioassay stations were mispredicted.

1986 AET values (dry weight normalization) are less efficient but are of similar sensitivity to the 1988 AET (dry weight normalization) and so have somewhat lower overall reliability. The sensitivity of 1988 amphipod bioassay AET is actually slightly higher than that for 1986 AET, although AET have increased for many chemicals (e.g., PAH). This net increase in sensitivity rather than a decrease is explained by the following four factors:

1. Increasing an AET value does not necessarily mean that previously predicted stations will no longer be predicted, because the concentration of chemicals at such stations may still exceed the updated AET value. Alternatively, the impacted stations may still be predicted by AET for other chemicals that are also detected at elevated concentration at the station.

2. Implementation of consistent rules for treatment of anomalous biological and chemical data in developing the 1988 AET resulted in the lowering of AET values relative to corresponding 1986 AET for five chemicals used in predictions.

3. 1988 AET are available for some chemicals that either were not used in predictions by 1986 AET (e.g., antimony, chromium) or were not analyzed for in the surveys used to establish 1986 AET (e.g., new data are available for resin acids and chlorinated phenols in the Everett Harbor survey and selected tentatively identified compounds detected in the Everett Harbor and Elliott Bay surveys).

4. Some 1986 AET are "greater than" values (i.e., preliminary AET that require further confirmation) for which concentrations of the chemical at all impacted stations are less than the highest concentration at a nonimpacted station in the surveys used to establish 1986 AET. These preliminary AET are not used in determining sensitivity. After addition of new survey data, 1988 AET for some of these chemicals became defined AET because at least one impacted station in the new surveys exceeded the highest concentration at a nonimpacted station. Such AET are used in determining sensitivity.

Of the 201 benthic infauna stations and 287 amphipod bioassay stations evaluated, approximately 75 percent (157 stations and 213 stations, respectively) are in accordance with the predictions of the 1986 AET values for these indicators. AET for the oyster larvae and Microtox bioassays have not been updated because new data using comparable bioassay protocols were not available (data are only available for Commencement Bay for these two indicators). Approximately 96 percent of both the existing 50-station oyster larvae database and the 50-station Microtox database are in accordance with the predictions of the 1986 values for these indicators.

TOC-normalized AET did not appear to offer an advantage in predictive success over dry weight-normalized AET. Specifically, for 1988 AET calculated for amphipod and benthic infauna indicators, use of TOC-normalized AET for nonionic organic compounds and dry weight-normalized AET for the remaining chemicals (metals and ionizable organic compounds) yielded approximately equivalent reliability results when compared with predictions by dry weight-normalized AET for all chemicals. Different hypotheses are discussed in the report for why TOC-normalized AET may be no more predictive than dry-weight AET.

Chemical and biological data for impacted stations that were not predicted as impacted by AET were reviewed to assess possible reasons for the incorrect predictions including the following:

- The indicated adverse biological effect resulted from factors other than toxic chemical exposures (e.g., physical disruption, grain size distribution)

- Unidentified chemicals accounted for the sediment toxicity

- High chemical detection limits precluded an assessment of whether the AET was exceeded

- The biological effect was incorrectly classified as statistically significant.

Physical characteristics (e.g., grain size, habitat exposure) and high detection limits (for which AET predictions are not made), or unanalyzed chemicals are the most likely factors accounting for the 29 amphipod and 24 benthic infauna stations that were not in accordance with predictions based on AET values. High detection limits for at least one chemical at almost all of these stations prevented a complete comparison to AET values. A relatively high percentage (≥75 percent) of fine-grained sediment was found at 55 percent of the incorrectly predicted amphipod bioassay stations (81 percent of the incorrectly predicted stations in the EIGHTBAY survey); such conditions are less favorable for the test organisms. A significant depression in benthic infaunal abundance was found at one of the two fine-grained sediments at which data for both biological indicators were available. These potential physical factors increase the uncertainty with which adverse biological effects can be identified. However, for the existing database, the 1988 AET are efficient with respect to not predicting effects for stations at which grain size or related factors may dominate over toxic effects.

A relatively high percentage (≥70 percent) of coarse-grained sediment was found at 64 percent of the incorrectly predicted benthic infauna stations and in some cases these samples were collected in areas of relatively high-wave intensity compared to reference conditions. Both conditions are expected to result in reduced abundances of organisms that may not have been adequately controlled for by comparisons to reference area samples. Significant mortality was observed in the amphipod bioassay at two of the eight coarse-grained sediment stations at which data for both biological indicators were available.

Detection limits, unmeasured chemicals, or unaccounted chemical interactions are important considerations at the five stations at which both significant amphipod mortality and significant depressions in abundance of benthic infauna occurred but no AET were exceeded. These stations are of concern because AET for both indicators failed to predict the observed effects and grain-size conditions were generally not considered to be a factor (these five stations constitute 3 percent of the 154 stations at which both amphipod bioassay and benthic infaunal data are available for comparisons). The five stations are from the PSEP Elliott Bay survey, and four of the five are located near Harbor Island in the lower Duwamish River, a federal Superfund site contaminated by a variety of organic and inorganic chemicals. The fifth station is located near Pier 91 in Elliott Bay; untested but potential tributyltin contamination from historical naval ship operations should be considered as a possible explanation for the substantial adverse effects observed at this station.

In general, there was little evidence to suggest that the reliability of AET was affected by geographic location. The 1988 AET are 100 percent efficient in predicting effects in areas of Puget Sound represented in the database (excluding chemically anomalous stations at which relatively high chemical concentrations but no adverse effects were observed). After including chemical anomalies in predictions, the 1988 amphipod AET are 88 percent efficient and the 1988 benthic infauna AET are 96 percent efficient for a total database of 295 and 205 stations, respectively. High and nearly identical sensitivity was observed for both amphipod bioassay and benthic infaunal abundance indicators in the Commencement Bay and Everett Harbor surveys conducted over two years apart in north and south Puget Sound. In general there were greater differences in sensitivity among individual surveys in the same location than among geographic areas. The lowest sensitivity (e.g., 17 percent for amphipod bioassay results in the EIGHTBAY study) was associated with surveys for which potential grain-size effects, high detection limits, or lack of data for all chemical classes were major factors.

From 60-100 percent sensitivity was observed in all geographic areas for which all chemical classes represented by AET were analyzed in a survey (including tentatively identified organic compounds). In a separate test, AET were generated solely from upper Duwamish River stations sampled in two surveys and compared with a later independent survey in the same area as well as with data from Everett Harbor and Commencement Bay. High (100 percent) sensitivity and low (31-35 percent) efficiency was attained in each of these comparisons. Hence, there was no evidence of differences in reliability among geographic regions although the test results indicated that a sample size of 16 stations from one geographic area was too small or nonrepresentative of a range of contaminant conditions to generate efficient AET values.

Based on these results and further tests using random selection of stations from the entire database, it is recommended that the following strategy be used to develop reliable AET: 1) Collect chemical and biological effects data from preferably more than 50 stations; 2) Bias the positioning of stations to ensure sampling of a variety of contaminant sources (e.g., an urban environment impacted by multiple contaminant sources, and preferably representative of the range of geographic areas to which the AET will be applied) over a range of contaminant concentrations (preferably over at least one to two orders of magnitude); 3) Conduct chemical tests for a wide range of chemical classes and ensure that <100 ppb detection limits (lower if possible) are attained for organic compounds (metals detection limits do not appear to be a problem). In general, small numbers of randomly selected stations yield AET that are sensitive (>80-100 percent) but not highly efficient (40-60 percent).

The sensitivity of 1988 AET is attributable to several chemical classes. No one chemical or chemical class (e.g., PAH, metals, phenolic compounds) accounts for more than half of the sensitivity that is attained using AET for all chemicals. However, a number of chemicals do not contribute to the sensitivity of the 1988 AET because they are infrequently detected or reported in Puget Sound (Table 14). In addition, several chemical variables do not appear to uniquely account for predicted impacts for either the amphipod bioassay or benthic infauna indicators, including standard conventional variables (TOC, TVS, sulfides, grain size, oil & grease), volatile organic compounds (except potentially trichloroethene in the Duwamish River), and several miscellaneous compounds (e.g., pentachlorophenol, resin acids, organic bases, dibenzofuran, benzoic acid, benzyl alcohol). Routine analysis for volatile organic compounds may not be warranted based on these results, which would represent a cost savings because these compounds require separate analysis.

The chemical groups that contribute most substantially to the sensitivity of the 1988 AET for amphipod bioassay and benthic infauna indicators include PAH, metals (excluding beryllium, selenium, barium, and naturally occurring iron or manganese), phenols, phthalate esters (not necessarily the few phthalate esters that are common laboratory contaminants), chlorinated neutral compounds (e.g., chlorinated benzenes, butadienes, and PCBs), and tentatively identified compounds (including PAH-related compounds and miscellaneous oxygenated compounds). An evaluation of the incremental increase in sensitivity by adding these chemical groups (Figure 9) indicated that PAH and metals were the two most important for identifying impacts in the Puget Sound database. However, these groups are also two of the most commonly measured variables, which may bias the results against other compound groups that have not been as widely measured at low detection limits (e.g., alkyl- and chlorinated phenols or chlorinated benzenes).

Finally, based on AET for the four available biological indicators (amphipod, oyster larvae, and Microtox bioassays, and benthic infaunal abundance), maps were

prepared showing the location of predicted impacts in Puget Sound using most of the 334-station database. For ranking these stations, both the magnitude of AET exceedance and the number of AET exceedances are important in establishing a preponderance of evidence for predicting adverse biological effects. Comparison of the cumulative ratio of the concentration of each sediment contaminant to its AET value was suggested as a possible way to rank these stations and conform to an additive model of chemical toxicity.

CONTENTS

FIGURES

APPENDICES

TABLES

APPENDICES

ACKNOWLEDGEMENTS

This document was prepared by PTI Environmental Services under the direction of Mr. Robert Barrick for the U.S. Environmental Protection Agency (EPA) in partial fulfillment of Contract No. 68-02-4341. Funding for the project was provided by the National Estuary Program, Office of Marine and Estuarine Protection. Ms. Catherine Krueger served as technical monitor for EPA Region 10 and the Puget Sound Estuary Program. Mr. Keith Phillips of the Washington Department of Ecology also provided direction in initially defining the scope of this project and in developing the conceptual framework for sediment management applications of reliability analyses. Technical review and constructive comments on draft reports and the presentation of concepts were provided by individual members of the Puget Sound Sediment Criteria Work Group.

The primary authors of this report are Mr. Robert Barrick, Dr. Scott Becker, Ms. Lorraine Brown, Mr. Harry Beller, and Dr. Robert Pastorok. Additional technical and editorial review was provided by Dr. Thomas Ginn and Ms. Carol Newlin. Mr. Dreas Nielsen and Dr. Chip Hogue developed SEDQUAL, the menu-driven database system used for conducting the extensive comparisons of sediment quality values provided in this report. Sponsoring agencies provided substantial support in defining the overall management context within which development of sediment quality values must occur.

1.0 INTRODUCTION

This report consists of updated evaluations of the predictive capability of Apparent Effect Threshold (AET) values, used as one approach to develop sediment quality values for Puget Sound (Beller et al. 1986). The application of AET to predict biological effects in an expanded chemical database for Puget Sound is also presented. The work was conducted as part of the Sediment Quality Values Refinement project, which is sponsored by the Puget Sound Estuary Program (PSEP) through the Office of Puget Sound. Initiation of the study was approved by the Puget Sound Implementation Committee and Technical Advisory Committee. The results of this study are intended for use by PSEP and the Puget Sound Dredged Disposal Analysis (PSDDA) program, the Washington Department of Ecology, and the Puget Sound Water Quality Authority in defining guidelines for managing contaminated sediments in Puget Sound. This report does not address policy decisions for use of AET by agencies. It is also recommended that the AET values not be applied outside Puget Sound without further evaluations.

Methods used to develop and evaluate AET values are summarized in Section 2. Results of the evaluations, which include validation tests using Puget Sound data, are presented in Section 3 and summarized in Section 4. References and a glossary of terms are provided in Sections 5 and 6, respectively. Written comments received by the sponsoring agencies from reviewers of the draft report have been incorporated in this final report whenever possible.

1.1 BACKGROUND

During 1986 and 1987, the Office of Puget Sound, U.S. EPA Region 10, the U.S. Army Corps of Engineers Seattle District, and the Washington Departments of Ecology and Natural Resources undertook a joint effort to develop sediment quality values for Puget Sound. This effort was funded by PSEP and PSDDA. The goal of the project was to identify concentrations of chemicals in sediments that are expected (based on field evidence or theoretical predictions) to be associated with adverse biological effects. The specific objectives of the study were to compile synoptic biological and chemical data from Puget Sound, to evaluate techniques that could be used to develop chemical-specific sediment quality values, and to evaluate the reliability (i.e., ability to correctly identify sites with known biological effects) of the values generated using different techniques. The study was completed in September 1986 and copies of the final report (Beller et al. 1986) were distributed nationally and in the Puget Sound region.

The 1986 report contains an evaluation of several different approaches to establishing sediment quality values and ranges of chemical values that can be applied in sediment management, but does not identify the specific methods or values that the agencies will adopt for regulation or how these numerical values should be modified for application in specific regulatory programs.

Based on the conclusions of the PSEP/PSDDA study, the agencies requested a more detailed evaluation of the use of the AET approach in establishing sediment quality values. An AET is defined as the sediment concentration of a chemical above which a particular adverse biological effect is always expected. AET are considered distinct from sediment quality values. AET are tools that can be used in sediment

management. Sediment quality values, which may be based in part on AET, have yet to be adopted by PSEP or PSDDA and, as a policy decision, may include safety factors or other modifications.

1.2 OBJECTIVES

The overall objective of Work Assignment No. 10 is to further test the reliability of AET, and potential resulting sediment quality values based on AET in predicting adverse biological effects, and to provide more detailed analyses of the relationships between chemical concentrations and biological effects. Most of these evaluations were conducted as part of the 1988 recalculation and evaluation of Puget Sound AET presented in this report (Volume I). The relationship of these tasks to other tasks performed under the sediment quality values work assignment (and presented in Barrick et al. 1988; Volume II) is shown in Figure 1.

The following six objectives were accomplished as tests of the reliability of AET:

1. Further test the predictive reliability of 1986 Puget Sound AET using new data sets for Eagle Harbor, Elliott Bay, and Everett Harbor (see Section 3.1)

2. Evaluate the predictive reliability of Puget Sound AET after incorporation of these new data sets in the calculation of 1988 AET (see Section 3.2)

3. Investigate the characteristics of and develop hypotheses for stations at which observed biological effects run counter to AET predictions (see Section 3.3.1)

4. Test the influence of geographic location on AET reliability (see Section 3.3.2)

5. Determine if a subset of chemicals in the sediment quality database accounts for all or nearly all of the predictive reliability of AET (see Section 3.3.4).

6. Compare the predictive reliability of the 1986 and 1988 Puget Sound AET using independent PSDDA data sets (e.g., Oak Harbor, Port Gardner, and Duwamish River studies) [note: only a limited comparison was possible because of incomplete chemical or biological test results for most of these dredging studies; see Appendix G].

The following three additional objectives were accomplished by applying the sediment quality values database to Puget Sound data:

1. Investigate minimum data requirements to derive reliable AET (see Section 3.3.3)

2. Identify potential problem areas by predicting biological effects in Puget Sound from chemical data (see Section 3.4)

3. Evaluate the influence of different bioassay reference data sets on the development of AET (see Appendix C).

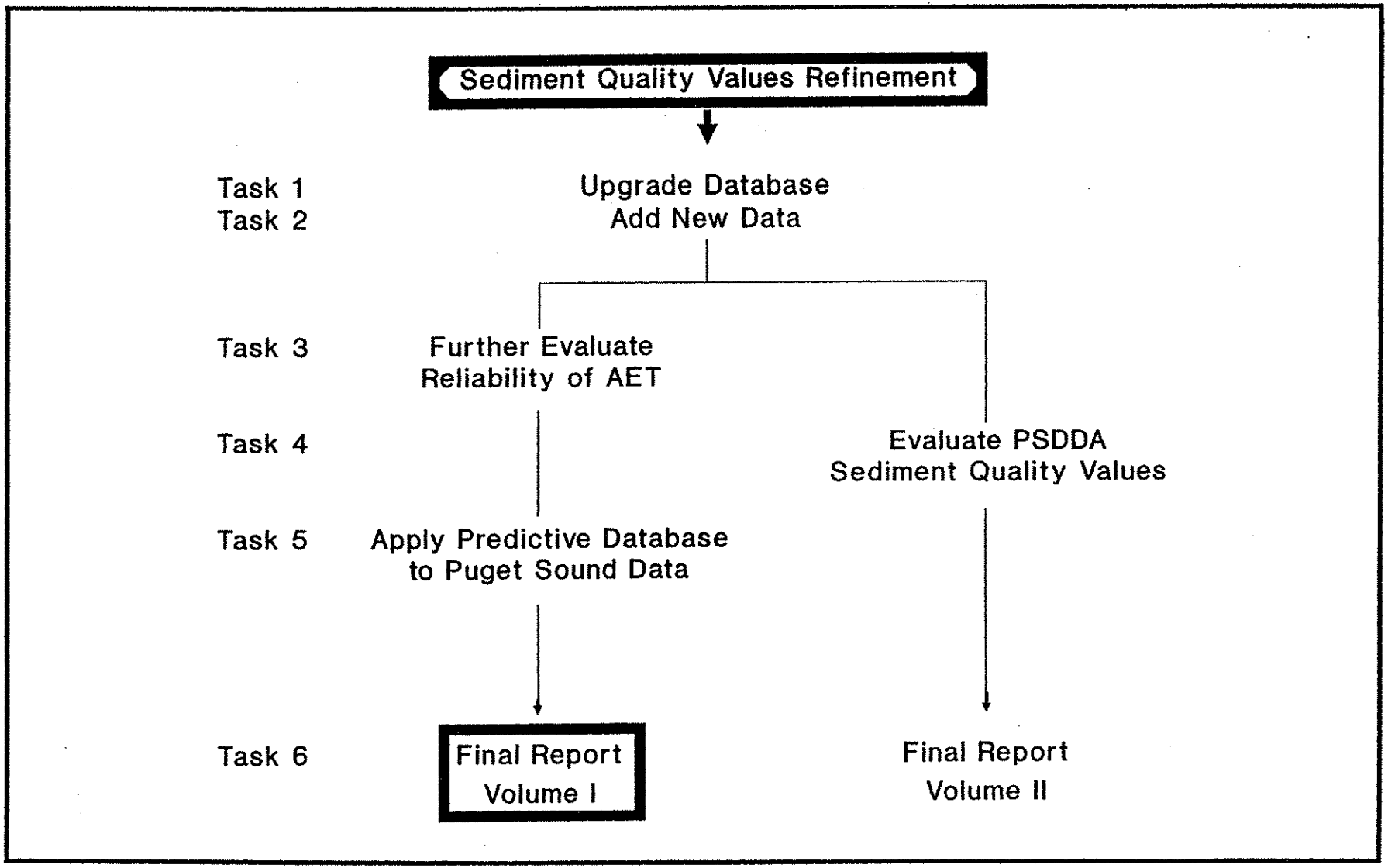


Figure 1. Relationship of sediment quality values refinement tasks.

2.0 METHODS

Most analyses for this report were conducted using a menu-driven program (SEDQUAL) developed for the Puget Sound sediment quality values database. A detailed description of the database and menu-driven features is given in the SEDQUAL users manual (Nielsen 1988). In this section, a summary is provided of the AET approach, measures of reliability (i.e., sensitivity and efficiency of the predictions of adverse biological effects based on chemical data), and the data sets used to test and expand the AET database. Biological analyses and statistical procedures used to determine adverse biological effects have been summarized by Beller et al. (1986). Modifications to these procedures that were used for this report are discussed briefly in this section (see Appendix C for details).

To accomplish the objectives listed in Section 1.0 (INTRODUCTION), 1986 Puget Sound AET for individual biological indicators (amphipod bioassay, oyster larvae bioassay, Microtox bioassay, and benthic infaunal abundance) were compiled in SEDQUAL and used to compare against chemical concentration data for Eagle Harbor, Elliott Bay, and Everett Harbor sediments. Calculations were performed to determine the reliability of these AET in correctly identifying impacted and nonimpacted stations for these data sets. Lists were compiled for evaluation of those stations that were incorrectly predicted by these AET (i.e., either exhibiting an adverse biological effect when none was predicted, or not exhibiting an adverse biological effect when one was predicted). Iterative procedures were developed and performed to test the predictive efficiency and sensitivity of the AET database. Specific details of these procedures are discussed in Section 3.0 (RESULTS).

2.1 MEASURES OF RELIABILITY: SENSITIVITY AND EFFICIENCY

Environmental factors such as matrix effects or chemical interactions can affect the ability to correctly predict adverse biological effects in environmental samples based on any chemical-specific criteria. The only way to definitively confirm the chemical-specific predictions (i.e., whether particular biological effects are always observed above a given sediment quality value) is to conduct controlled laboratory spiking studies. A second approach is to test field samples with known concentrations of the contaminant of concern but this approach does not directly quantify possible interactions among chemicals. However, the binary (impacted/nonimpacted) predictions of sediment quality values, which are most pertinent to sediment management, are testable using data sets that have matched chemical and biological data. Tests using such data are applicable to and recommended for assessing environmental predictions of any approach to developing sediment criteria as part of management performance objectives. The basic requirements of such tests of reliability are described in this section.

To meet the needs of most sediment quality management programs, an ideal sediment criteria approach would perform well on both of the following measures of reliability, which are evaluated with actual field data:

- **Sensitivity** in detecting environmental problems (i.e., are <u>all</u> biologically impacted sediments identified by the predictions of the chemical sediment criteria?)

- **Efficiency** in screening environmental problems (i.e., are only biologically impacted sediments identified by the predictions of the chemical sediment criteria?).

As a measure of reliability, sensitivity is defined as the proportion of all stations exhibiting a particular adverse biological effect that are correctly predicted using sediment quality values for that biological indicator. Efficiency is defined as the proportion of all stations predicted to have a particular adverse biological effect that actually are impacted. The concepts of sensitivity and efficiency are illustrated in Figure 2.

Sensitivity and efficiency are independent measures of reliability. For example, a sediment criteria approach that sets values for a wide range of chemicals near their analytical detection limits will probably be sensitive but inefficient. That is, it will predict a large percentage of sediments with biological effects but will also predict many biologically unimpacted sediments with only slightly elevated chemical concentrations. Such an approach may be environmentally protective but also may result in overregulation that would not be cost effective.

Conversely, a sediment criteria approach that sets values at the upper end of the range of environmental concentrations may be efficient but insensitive. That is, a high percentage of the stations with predicted impacts may indeed be biologically impacted, but the approach may fail to predict other biologically impacted stations with moderate to high chemical concentrations. Such an approach may be cost-effective and defensible in pursuing high priority remedial actions (i.e., would not overregulate) but would not be environmentally protective.

The overall reliability of any sediment criteria approach addresses both components of sensitivity and efficiency. This measure is defined as the proportion of all stations for which correct predictions were made for either the presence or absence of adverse biological effects:

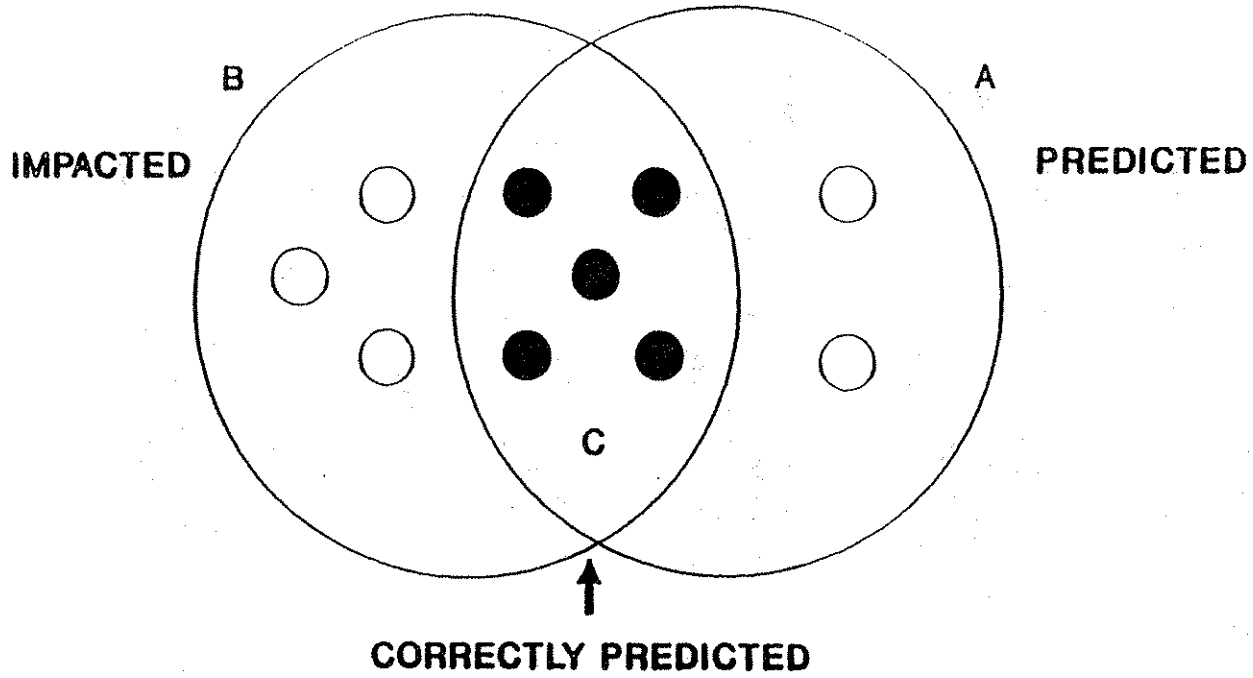
- **Overall reliability** = $\dfrac{\left[\begin{array}{c}\text{All stations correctly predicted as impacted} \\ + \\ \text{All stations correctly predicted as nonimpacted}\end{array}\right]}{\left[\text{Total number of stations evaluated}\right]}$

High reliability results from correct prediction of a large percentage of the impacted stations (i.e., high sensitivity; few false negatives) and correct prediction of a large percentage of the nonimpacted stations (i.e., high efficiency; few false positives). These measures of reliability are used as the primary means of evaluating AET values in this report.

2.2 SUMMARY OF AET APPROACH

Regulatory sediment criteria based on definitive laboratory cause-effect studies and field verification studies are still under development. In the interim, field effects-based approaches using the AET concept provide decision tools that have the following characteristics:

- Developed empirically from field data



SENSITIVITY = C/B x 100 = 5/8 x 100 = 63%

EFFICIENCY = C/A x 100 = 5/7 x 100 = 71%

FOR A GIVEN BIOLOGICAL INDICATOR:

A ALL STATIONS PREDICTED TO BE IMPACTED

B ALL STATIONS KNOWN TO BE IMPACTED

C ALL STATIONS CORRECTLY PREDICTED TO BE IMPACTED

Figure 2. Measures of reliability (sensitivity and efficiency).

- Expressed as chemical-specific values
- Supported by a variety of biological indicators, including lethal and sublethal effects on different life stages of organisms *in situ* or under laboratory conditions
- Driven by statistically significant ($P \leq 0.05$) adverse effects
- Supported by non-contradictory evidence of adverse effects within a given data set used to calculate AET.

AET can be developed for any measured chemical (organic or inorganic) that spans a wide concentration range in the data set used to generate AET. The AET concept can be applied to matched field data for sediment chemistry and any observable biological effects (e.g., bioassay responses, infaunal abundances at various taxonomic levels, bioaccumulation). By using these different biological indicators, application of the resulting sediment quality values enables a wide range of biological effects to be addressed in the management of contaminated sediments.

The focus of the AET approach is to identify concentrations of contaminants that are associated exclusively with sediments exhibiting statistically significant biological effects relative to reference sediments. The calculation of AET for each chemical and biological indicator is straightforward:

1. **Collect "matched" chemical and biological effects data**--Conduct chemical and biological effects testing on subsamples of the same field sample (to avoid unaccountable losses of benthic organisms, benthic infaunal and chemical analyses are conducted on separate samples collected concurrently for the same station)

2. **Identify "impacted" and "nonimpacted" stations**--Statistically test the significance of adverse biological effects relative to suitable reference conditions (i.e., responses resulting from exposure to sediments containing very low or undetectable concentrations of any toxic chemicals) for each sediment sample and biological indicator

3. **Identify AET using only "nonimpacted" stations**--For each chemical, the AET can be identified for a given biological indicator as the highest detected concentration among sediment samples that do not exhibit statistically significant effects (if the chemical is undetected in all these samples, no AET can be established for that chemical and biological indicator)

4. **Check for preliminary AET**--Verify that statistically significant biological effects are observed at a chemical concentration higher than the AET; otherwise the AET is only a preliminary minimum estimate (or may not exist).

5. Repeat Steps 1-4 for each biological indicator.

Additional discussion of the AET concept and a pictorial representation of the AET approach for two example chemicals are provided in Appendix A. Puget Sound AET values based on biological effects data for eight surveys were previously generated and evaluated by PSEP and PSDDA in 1986 (Beller et al. 1986). Selected 1986 AET for

Puget Sound data are summarized in Table 1 (dry weight normalization; a complete list of AET used in validation tests is provided in Appendix E). Additional evaluations of 1986 AET based on predictions using new data (e.g., surveys in Eagle Harbor, Elliott Bay, and Everett Harbor) are discussed in Section 3.1 of this report (RESULTS).

The 1986 Puget Sound AET for amphipod bioassays and benthic infauna analyses were recalculated as part of the current study by incorporating data from the Eagle Harbor, Elliott Bay, and Everett Harbor sampling programs, and applying guidelines for treating biological and chemical data used to develop sediment quality values (see Section 2.4 and Appendix C). 1988 AET for selected chemicals normalized to dry weight are provided in Table 2 (a complete list of 1988 AET used in validation tests is provided in Appendix F). For comparison, AET for the same chemicals normalized to total organic carbon (TOC) content are provided in Table 3. The chemicals presented in Tables 1-3 are among the most commonly measured and detected chemicals in Puget Sound sediments.

For comparison to AET values in Table 3, a summary of concentration ranges of selected chemicals and key sediment characteristics [i.e., total organic carbon (TOC), fine-grained sediment content, sulfides] is provided in Table 4 for biological effects stations contained in the Puget Sound sediment quality values database. Data for two groups of biological samples are given in Table 4: sediments sampled from reference areas (nonurban embayments) and from nonreference areas of Puget Sound (frequently from industrialized, urban embayments). Reference areas in Puget Sound are generally removed from the direct influence of contaminant sources in nonurbanized areas of the sound; almost all of these stations are nonimpacted. Most data for nonreference areas are from industrialized urban embayments and include both nonimpacted and impacted stations.

The predictive ability of the 1988 AET in identifying all impacted stations in the enhanced data set is evaluated in Section 3.2 (RESULTS). Evaluations of both the 1986 and 1988 AET focus on the ability of AET to correctly identify stations with adverse biological effects. As part of this evaluation, emphasis was placed on the sensitivity and efficiency of the predictions as two measures of reliability (see definitions in Section 2.1.

2.3 DATA SETS USED FOR ANALYSES

The 1986 Puget Sound AET are based on data collected in eight Puget Sound surveys (see Table B-1 in Appendix B and Figure 3). Additional data sets used in the current study to test the predictive capability of 1986 AET, or to generate 1988 AET values, include amphipod bioassay and/or benthic infaunal analyses conducted for studies at Eagle Harbor (8 stations for both biological indicators), Elliott Bay (72 benthic infauna stations and 92 amphipod bioassay stations), Everett Harbor (17 benthic infauna stations and 26 amphipod bioassay stations), and associated reference areas (9 benthic infauna stations and 9 amphipod bioassay stations in Blakely Harbor and Port Susan). Stations currently excluded as biological or chemical anomalies (see Section 2.4 and Appendix C) are not included in these totals.

The Eagle Harbor biological effects data were collected during a preliminary investigation conducted for Ecology's state Superfund program. PSEP sponsored the collection of bioassay and benthic infauna data for 10 stations sampled during the investigation, primarily to provide data for evaluation of sediment quality values at a site known to be heavily contaminated with polynuclear aromatic hydrocarbons (PAH). The Elliott Bay and Everett Harbor data were collected as part of Urban Bay Toxics Action Programs conducted by PSEP in these embayments located adjacent to the cites

TABLE 1. 1986 PUGET SOUND AET FOR SELECTED CHEMICALS (dry weight)[a]

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Metals (mg/kg dry weight; ppm)				
Antimony	5.3	26	3.2	26
Arsenic	93	700	85	700
Cadmium	6.7	9.6	5.8	9.6
Copper	810[f]	390	310	390
Lead	660	660	300	530
Mercury	2.1[f]	0.59	0.88	0.41
Nickel	>120[f]	39	49	28
Silver	>3.7[f]	>0.56	5.2	>0.56
Zinc	870[f]	1,600	260	1,600
Organic Compounds (ug/kg dry weight; ppb)				
Low molecular weight PAH	5,200	5,200	6,100[f]	5,200
Naphthalene	2,100	2,100	2,100	2,100
Acenaphthylene	560	>560	640[f]	>560
Acenaphthene	630[f]	500	500	500
Fluorene	540	540	640[f]	540
Phenanthrene	5,400[f]	1,500	3,200[f]	1,500
Anthracene	1,900[f]	960	1,300[f]	960
2-Methylnaphthalene	670	670	670	670
High molecular weight PAH	18,000[f]	17,000	>51,000[f]	12,000
Fluoranthene	3,900[f]	2,500	6,300[f]	1,700
Pyrene	4,300[f]	3,300	>7,300[f]	2,600
Benz(a)anthracene	1,600[f]	1,600	4,500[f]	1,300
Chrysene	2,800[f]	2,800	6,700[f]	1,400
Benzofluoranthenes	3,700	3,600	8,000[f]	3,200
Benzo(a)pyrene	2,400	1,600	6,800[f]	1,600
Indeno(1,2,3-c,d)pyrene	690[f]	690	>5,200[f]	600
Dibenzo(a,h)anthracene	260[f]	230	1,200[f]	230
Benzo(g,h,i)perylene	740[f]	720	5,400[f]	670
Chlorinated organic compounds				
1,3-Dichlorobenzene	>170	>170	>170	>170
1,4-Dichlorobenzene	260	120	120	110
1,2-Dichlorobenzene	>350	50	50	35
1,2,4-Trichlorobenzene	51	64	64	31
Hexachlorobenzene (HCB)	130	230	230	70
Total PCBs	2,500[f]	1,100	1,100	130
Phthalates				
Dimethyl phthalate	160	160	160	71
Diethyl phthalate	>73	>73	97	>48
Di-n-butyl phthalate	>5,100	1,400	>5,100	1,400

TABLE 1. (Continued)

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Phthalates (continued)				
Butyl benzyl phthalate	>470	>470	470	63
Bis(2-ethylhexyl)phthalate	>3,100	1,900	1,900	1,900
Di-n-octyl phthalate	>590	>420	>69,000	--
Phenols				
Phenol	670[f]	420	1,200	1,200
2-Methylphenol	63	63	>72	>72
4-Methylphenol	1,200	670	670	670
2,4-Dimethyl phenol	>50	29	29	29
Pentachlorophenol	>140	>140	>140	>140
Miscellaneous Extractables				
Benzyl alcohol	73	73	73	57
Benzoic acid	>690	650	650	650
Dibenzofuran	540	540	540	540
Hexachlorobutadiene	290	270	270	120
N-Nitrosodiphenylamine	220	130	74	40
Volatile Organics				
Tetrachloroethene	>210	140	140	140
Ethylbenzene	>50	37	37	33
Total xylenes	>160	120	120	100
Pesticides				
p,p'-DDE	15	--	9	--
p,p'-DDD	43	--	2	--
p,p'-DDT	3.9	>6	11	--

[a] ">" indicates that a defined AET could not be established because there were no "effects" stations with chemical concentrations above the highest concentration among "no effects" stations.

[b] Based on 150 stations.

[c] Based on 56 stations (all from Commencement Bay Remedial Investigation and Blair Waterway dredging study).

[d] Based on 94 stations.

[e] Based on 50 stations (all from Commencement Bay Remedial Investigation).

[f] The value shown exceeds AET established from Commencement Bay Remedial Investigation data (Barrick et al. 1985) because of addition of Puget Sound data presented in Beller et al. (1986).

TABLE 2. 1988 PUGET SOUND AET
FOR SELECTED CHEMICALS (normalized to dry weight)[a]

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Metals (mg/kg dry weight; ppm)				
Antimony	200[g]	--	150[g]	--
Arsenic	93	700	57[h]	700
Cadmium	6.7	9.6	5.1[h]	9.6
Chromium	270[g]	--	260[g]	--
Copper	1300[fg]	390	530[g]	390
Lead	660	660	450[g]	530
Mercury	2.1[f]	0.59	2.1[g]	0.41
Nickel	>140[fg]	--	>140[g]	--
Silver	6.1[fg]	>0.56	>6.1[g]	>0.56
Zinc	960[fg]	1,600	410[g]	1,600
Organic Compounds (ug/kg dry weight; ppb)				
Low molecular weight PAH	24,000[g]	5,200	13,000[fg]	5,200
Naphthalene	2,400[g]	2,100	2,700[g]	2,100
Acenaphthylene	1,300[g]	>560	1,300[fg]	>560
Acenaphthene	2,000[fg]	500	730[g]	500
Fluorene	3,600[g]	540	1,000[fg]	540
Phenanthrene	6,900[fg]	1,500	5,400[fg]	1,500
Anthracene	13,000[fg]	960	4,400[fg]	960
2-Methylnaphthalene	1,900[g]	670	1,400[g]	670
High molecular weight PAH	69,000[fg]	17,000	69,000[fg]	12,000
Fluoranthene	30,000[fg]	2,500	24,000[fg]	1,700
Pyrene	16,000[fg]	3,300	16,000[fg]	2,600
Benz(a)anthracene	5,100[fg]	1,600	5,100[fg]	1,300
Chrysene	9,200[fg]	2,800	9,200[fg]	1,400
Benzofluoranthenes	7,800[g]	3,600	9,900[fg]	3,200
Benzo(a)pyrene	3,000[g]	1,600	3,600[fh]	1,600
Indeno(1,2,3-c,d)pyrene	1,800[fg]	690	2,600[fh]	600
Dibenzo(a,h)anthracene	540[fg]	230	970[fh]	230
Benzo(g,h,i)perylene	1,400[fg]	720	2,600[fh]	670
Chlorinated organic compounds				
1,3-Dichlorobenzene	>170	>170	>170	>170
1,4-Dichlorobenzene	120[h]	120	110[h]	110
1,2-Dichlorobenzene	>110[h]	50	50	35
1,2,4-Trichlorobenzene	51	64		31
Hexachlorobenzene (HCB)	130	230	22[h]	70
Total PCBs	3,100[fg]	1,100	1,000[h]	130

TABLE 2. (Continued)

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Phthalates				
Dimethyl phthalate	>1,400[g]	160	>1,400[g]	71
Diethyl phthalate	>1,200[g]	>73	200[g]	>48
Di-n-butyl phthalate	1,400[h]	1,400	>5,100	1,400
Butyl benzyl phthalate	900[g]	>470	900[g]	63
Bis(2-ethylhexyl)phthalate	>3,100	1,900	1,300[h]	1,900
Di-n-octyl phthalate	>2,100[g]	>420	6,200[h]	--
Phenols				
Phenol	1,200[fg]	420	1,200	1,200
2-Methylphenol	63	63	72[g]	>72
4-Methylphenol	3,600[g]	670	1,800[g]	670
2,4-Dimethyl phenol	72[g]	29	210[g]	29
Pentachlorophenol	360[g]	>140	690[g]	>140
Miscellaneous Extractables				
Benzyl alcohol	870[g]	73	870[g]	57
Benzoic acid	760[g]	650	650	650
Dibenzofuran	1,700[g]	540	700[g]	540
Hexachlorobutadiene	180[h]	270	11[h]	120
N-Nitrosodiphenylamine	48[h]	130	28[h]	40
Volatile Organics				
Tetrachloroethene	>210	140	57[h]	140
Ethylbenzene	>50	37	10[h]	33
Total xylenes	>160	120	40[h]	100
Pesticides				
p,p'-DDE	15	--	9	--
p,p'-DDD	43	--	16[g]	--
p,p'-DDT	>270[g]	>6	34[g]	--

[a] ">" indicates that a defined AET could not be established because there were no "effects" stations with chemical concentrations above the highest concentration among "no effects" stations. "--" indicates AET data not available.

[b] Based on 287 stations (including recent surveys in Eagle Harbor, Elliott Bay, and Everett Harbor not included in the previous generation of 1986 AET).

[c] Based on 56 stations (all from Commencement Bay Remedial Investigation and Blair Waterway dredging study); unchanged since 1986.

TABLE 2. (Continued)

[d] Based on 201 stations (updated from Table 1 by incorporation of recent surveys in Eagle Harbor, Elliott Bay, and Everett Harbor not included in the previous generation of 1986 AET).

[e] Based on 50 stations (all from Commencement Bay Remedial Investigation); unchanged from Table 1.

[f] The value shown exceeds AET established from Commencement Bay Remedial Investigation data (Barrick et al. 1985) because of addition of Puget Sound data presented in Beller et al. (1986).

[g] The value shown exceeds AET presented in Beller et al. (1986) because of addition of Puget Sound data from the Eagle Harbor, Elliott Bay, or Everett Harbor surveys.

[h] The value shown is less than AET presented in Beller et al. (1986) because of the exclusion of chemically or biologically anomalous stations (see Table 6 and Appendix C) from the AET dataset.

TABLE 3. 1988 PUGET SOUND AET FOR SELECTED CHEMICALS (normalized to total organic carbon)[a]

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Nonionic Organic Compounds (mg/kg organic carbon; ppm)				
Low molecular weight PAH	2,200	370	780	>530
Naphthalene	220	99	170	>170
Acenaphthylene	66	>27	66	>27
Acenaphthene	200	16	57	>57
Fluorene	360	23	79	>71
Phenanthrene	690	120	480	>160
Anthracene	1,200	>79	220	>79
2-Methylnaphthalene	>120	---	64	---
High molecular weight PAH	5,300	960	7,600	1,500
Fluoranthene	3,000	160	1,200	>190
Pyrene	1,000	>210	1,400	>210
Benz(a)anthracene	270	110	650	>160
Chrysene	460	110	850	>200
Benzofluoranthenes	450	230	1,500	>430
Benzo(a)pyrene	210	99	>1,300	>140
Indeno(1,2,3-c,d)pyrene	88	33	900	>87
Dibenzo(a,h)anthracene	47	120	89	33
Benzo(g,h,i)perylene	78	31	>1,200	>67
Chlorinated benzenes				
1,3-Dichlorobenzene	>15	>15	>15	>15
1,4-Dichlorobenzene	9	3.1	16	>16
1,2-Dichlorobenzene	>5.8	2.3	2.3	2.3
1,2,4-Trichlorobenzene	1.8	2.7	---	0.81
Hexachlorobenzene (HCB)	4.5	9.6	0.38	2.3
Total PCBs	190	>46	65	12
Phthalates				
Dimethyl phthalate	53	>22	53	>19
Diethyl phthalate	>110	>5.3	61	>5.3
Di-n-butyl phthalate	260	260	1,700	220
Butyl benzyl phthalate	42	>9.2	64	4.9
Bis(2-ethylhexyl)phthalate	78	60	60	47
Di-n-octyl phthalate	58	>57	4,500	--

TABLE 3. (Continued)

Chemical	Amphipod AET[b]	Oyster AET[c]	Benthic AET[d]	Microtox AET[e]
Miscellaneous Extractables				
Dibenzofuran	>170	15	58	>58
Hexachlorobutadiene	6.2	11	6.9	3.9
N-nitrosodiphenylamine	>11	>11	11	>11
Volatile Organics				
Tetrachloroethene	>22	>22	>22	>22
Ethylbenzene	>3.8	>3.8	>3.8	>3.8
Total xylenes	>12	>12	>12	>12
Pesticides				
p,p'-DDE	0.81	--	0.31	--
p,p'-DDD	2.2	--	1.0	--
p,p'-DDT	>16	--	3.7	--
Ionizable Organic Compounds (mg/kg organic carbon; ppm)				
Phenols and Miscellaneous Extractables				
Phenol	440	>39	>140	33
2-Methylphenol	3.1	3.1	10	>10
4-Methylphenol	780	37	250	81
2,4-Dimethyl phenol	6.5	>1.3	2.6	0.63
Pentachlorophenol	24	>11	66	>11
Benzyl alcohol	73	5.0	>73	5.0
Benzoic acid	>170	>170	>170	>170
Metals (mg/kg organic carbon; ppm)				
Antimony	>55,000	3,300	5,500	3,300
Arsenic	32,000	88,000	4,400	88,000
Cadmium	1,100	1,200	580	1,200
Chromium	>150,000	---	65,000	---
Copper	100,000	49,000	13,000	48,000
Lead	110,000	66,000	18,000	66,000
Mercury	210	210	120	77
Nickel	>41,000	---	31,000	---
Silver	170	>100	490	100
Zinc	210,000	>200,000	48,000	>200,000

TABLE 3. (Continued)

[a] ">" indicates that a defined AET could not be established because there were no "effects" stations with chemical concentrations above the highest concentration among "no effects" stations (normalized to TOC). "--" indicates AET data not available.

[b] Based on 287 stations (including recent surveys in Eagle Harbor, Elliott Bay, and Everett Harbor not included in the previous generation of 1986 AET).

[c] Based on 56 stations (all from Commencement Bay Remedial Investigation and Blair Waterway dredging study); unchanged since 1986.

[d] Based on 201 stations (including recent surveys in Eagle Harbor, Elliott Bay, and Everett Harbor not included in the previous generation of 1986 AET).

[e] Based on 50 stations (all from Commencement Bay Remedial Investigation); unchanged since 1986.

TABLE 4. SUMMARY OF SELECTED CHEMICAL CONCENTRATIONS AND KEY SEDIMENT CHARACTERISTICS FOR BIOLOGICAL STATIONS FROM REFERENCE AND NONREFERENCE AREAS OF PUGET SOUND[a]

Chemical	Reference Area[b]		Non-Reference Area[c]	
	Minimum	Maximum	Minimum	Maximum
Metals (mg/kg dry weight; ppm)				
Antimony	0.13	2.86	0.1	1,370
Arsenic	1.9	22	0.34	9,700
Cadmium	0.047	1.9	0.3	184
Chromium	9.6	255	5.4	555
Copper	4.9	74	3.6	11,400
Lead	0.4	23	0.79	71,100
Mercury	0.016	0.134	0.06	52
Nickel	11	141	6.9	118
Silver	0.02	0.78	0.013	8.27
Zinc	15	102	18.4	6,010
Organic Compounds (ug/kg dry weight; ppb)				
Low molecular weight PAH	2.5	55	1.3	630,000
High molecular weight PAH	22	140	7.5	3,200,000
1,4-Dichlorobenzene	U5	U160	1	31,000
Hexachlorobenzene (HCB)	U0.3	U800	0.05	730
Hexachlorobutadiene	U1.3	U820	1	730
N-Nitrosodiphenylamine	U5	U2,500	4	610
4-Methylphenol	2	290	3	100,000
Phenol	13	560	0.9	2,900
Total PCBs	2.7	37	0.5	9,600
Conventional Sediment Variables (dry weight)				
Total organic carbon (%)	0.19	2.69	0.06	29.4
Fine-grained material (%)	7.4	88	2.7	95.5
Sulfides (mg/kg; ppm)	2.2	31	1	7,610

[a] Ranges are based only on biological stations used in the evaluation of Puget Sound AET. Higher concentrations for some chemicals have been detected in sediment samples for which there are no biological data, including several samples collected at depth. "U" indicates undetected in all samples over the detection limit range shown; otherwise the minimum value is the lowest detected concentration and the maximum value is the highest detected concentration (detection frequencies are not indicated but are less that 100 percent for most chemicals except metals).

[b] Reference areas in Puget Sound are generally removed from the direct influence of contaminant sources in nonurbanized areas of the sound; almost all of these stations are nonimpacted.

[c] Most data for nonreference areas are from industrialized urban embayments; nonreference data include **both** nonimpacted and impacted stations.

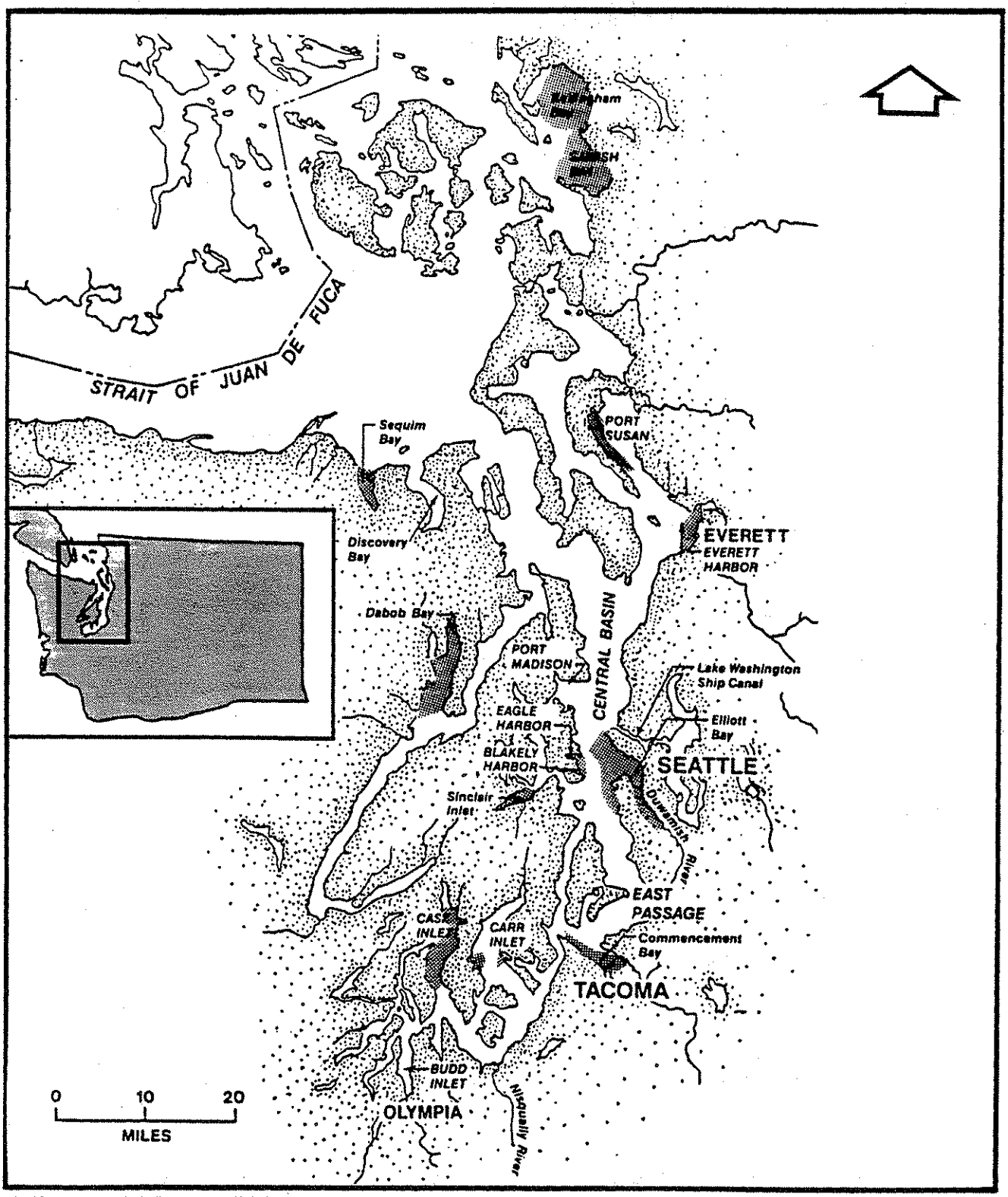


Figure 3. Location of sampling sites for Apparent Effects Threshold data sets in Puget Sound.

of Seattle and Everett, respectively. Contamination in each area is attributed to a variety of industrial, commercial, and residential sources.

A summary of stations and biological test types for all surveys that were used in this project is provided in Table 5. A summary of data sources is provided in Appendix B. Location maps of all stations are given in Figures B-1 through B-20. Data tables produced from SEDQUAL for comparisons to 1986 and 1988 AET are provided in Appendices E and F (see Volume I-Data Appendices). Although the AET approach can be applied to any set of matched chemical and biological data, the comparisons in this report are constrained to the biological effects data that are currently available for Puget Sound (i.e., primarily acute biological effects). Generation of AET in the future for biological indicators that address other adverse conditions (e.g., chronic sublethal effects) is possible.

Antimony, chromium, and nickel data are only considered in the evaluations of 1988 AET. Modifications of analytical procedures were instituted in PSEP protocols to better define the total concentration of these and other metals in sediments. These changes in analytical procedures were implemented in the Elliott Bay and Everett Harbor surveys conducted by PSEP (including reference area data collected in Port Susan). Additional tests conducted on sediment samples from Elliott Bay and Port Susan, and on archived sediment samples from Carr Inlet (a reference area in Southern Puget Sound) indicate that antimony and chromium concentrations are substantially higher with this revised technique than determined with the previous technique. As a result, 1986 Puget Sound AET values for antimony and chromium that are based on historical data by other metals analysis techniques are inefficient predictors of problem sediments in the new surveys. Definitive test results for X-ray fluorescence analyses also indicated that antimony measurements by other techniques (i.e., strong or total acid digestion) are less reliable than those for other metals. All but five test samples for antimony have been analyzed by these other techniques; hence, antimony AET values should be considered estimates only.

Nickel concentrations were comparable using the two metals digestion techniques. However, the range of nickel concentrations in the Puget Sound database prior to the addition of Elliott Bay and Everett Harbor data was narrow relative to Puget Sound reference concentrations. Thus, the 1986 AET for nickel was inefficient in reliably predicting adverse effects in the new surveys, which had a wider range of nickel concentrations. The 1988 AET for nickel are based on amphipod bioassay and benthic infaunal analyses that include Elliott Bay and Everett Harbor surveys.

2.4 TREATMENT OF DATA USED TO GENERATE 1988 AET

Before conducting analyses presented in Section 3.0 (RESULTS), options were developed to address statistical treatment of biological data, chemical data qualifiers, and anomalous chemical concentrations in data sets used to establish or verify sediment quality values. A summary of options and recommendations for treating anomalous data is presented in Appendix C. Such anomalous values are not necessarily incorrect or unreliable, but may be unrepresentative of Puget Sound conditions. Of special concern are stations that do not appear to have biological effects, but are chemically contaminated at concentrations that are well above those at other nonimpacted stations. The inclusion of such anomalous stations may increase AET values as a result of highly localized, nonrepresentative conditions. Alternatively, such stations may simply indicate the need for additional data to confirm an increase in AET values that is more representative of Puget Sound conditions.

TABLE 5. SUMMARY OF DATA SETS USED TO EVALUATE PUGET SOUND AET

Embayment	Survey Code	Number/Kind of Bioeffect Samples[a]	Chemical Analyses Conducted[b]							
			Acid	Base	Neut.	PCB	Pest.	VOA	Metal	Misc
Bellingham Bay	EIGHTBAY	8 / A	x	x	x	x	x	x	x	
Carr Inlet	CBMSQS	4 / BAOM	x	x	x	x	x	x	x	x
Case Inlet	EIGHTBAY	4 / A	x	x	x	x	x	x	x	
Central Puget Sound Basin	ALKI	4 / B	x	x	x	x	x	x	x	
	EHCHEM	2 / BA	x	x	x	x	x	x	x	x
Commencement Bay	CBBLAIR	6 / BAO	x	x	x	x	x	x	x	x
	CBMSQS	42 / BAOM	x	x	x	x	x	x	x	x
		2 / AOM								
		2 / B OM								
Dabob Bay	EIGHTBAY	4 / A	x	x	x	x	x	x	x	
Eagle Harbor	EHCHEM	8 / BA	x	x	x	x	x	x	x	x
Elliott Bay	EBCHEM	71 / BA	x	x	x	x	x	x	x	x
		24 / A								
		4 / B								
	ALKI	7 / B	x	x	x	x	x	x	x	
	TPPS3AB	27 / B	x	x	x	x	x	x	x	
	DUWRIV1	8 / A			x	x	x	x	x	
	DUWRIV2	30 / A			x	x	x	x	x	
	EIGHTBAY	8 / A	x	x	x	x	x	x	x	
Everett Harbor	EVCHEM	13 / BA	x	x	x	x	x	x	x	x
		13 / A								
		3 / B								
	EVERETT1	6 / A		x	x		x	x		
	EIGHTBAY	8 / A	x	x	x	x	x	x	x	
Port Susan	EBCHEM	5 / BA	x	x	x	x	x	x	x	x
	EVCHEM	3 / BA	x	x	x	x	x	x	x	x
Samish Bay	EIGHTBAY	4 / A	x	x	x	x	x	x	x	
Sequim Bay	EIGHTBAY	4 / A	x	x	x	x	x	x	x	
	DUWRIV1	1 / A			x	x	x	x	x	
	DUWRIV2	1 / A			x	x	x	x	x	
Sinclair Inlet	EIGHTBAY	8 / A	x	x	x	x	x	x	x	

TABLE 5. (Continued)

[a] 334 distinct samples (including 12 repeated samplings) at a total of 322 locations:
(B) 201 benthic infaunal analyses; (A) 287 amphipod mortality bioassays; (O) 56 oyster larvae abnormality bioassays; (M) 50 Microtox (saline extract) bioassays. The seven amphipod bioassay stations excluded as biological anomalies, and four benthic infauna and eight amphipod bioassay stations excluded as chemical anomolies (see Section 2.4 and Appendix C) are not included in these totals. Station locations for each survey are summarized in Appendix B.

[b] Chemical analyses conducted for U.S. EPA priority pollutant acid, base, neutral, PCB, pesticide, and volatile organic compounds, metals, and miscellaneous compounds not recognized as EPA priority pollutants (e.g., resin acid compound data for the EVCHEM survey, and tentatively identified organic compounds).

Implementation of all of the procedures recommended in Appendix C to all AET data types was not within the present scope of work. Nevertheless, recommended methods were applied whenever possible (see Table C-1 in Appendix C for a summary of implemented options). The methods for treatment of biological and chemical data used to generate 1988 AET presented in this report are described in the following sections. Implementation of the following biological and chemical guidelines for anomalies resulted in a net increase of 8 percent in the sensitivity of 1988 amphipod AET (from 50 to 58 percent) and a net increase of 4 percent in the sensitivity of 1988 benthic infauna AET (from 71 to 75 percent), when used to predict impacts in the expanded database. Because all Puget Sound data have been used to generate these 1988 AET, their efficiency is by definition 100 percent for the Puget Sound database of biological effect stations that pass both biological and chemical guidelines for anomalies.

2.4.1 Bioassay Data

In generating 1988 AET, several modifications to procedures used previously (e.g., Beller et al. 1986) were made for the inclusion of new data for the amphipod mortality bioassay. The recommendations discussed in Appendix C were adopted to improve the consistency of the results among the various pooled studies. Similar modifications were not made for the oyster larvae abnormality bioassay and the Microtox bioassay because each of these indicators is represented by only a single study area in the SEDQUAL database.

The first modification addressed the level of significance at which pairwise comparisons between impacted and reference sites were judged significant. A level of $P \leq 0.05$ comparisonwise was consistently used for all comparisons between stations, instead of using an experimentwise error rate that would result in variable *alpha* levels for pairwise station comparisons depending on the sample size of different studies (Appendix C).

The second modification considered screening criteria and power analyses for the amphipod tests conducted in the Elliott Bay and Everett Harbor PSEP surveys (EBCHEM and EVCHEM), discussed in detail in Appendix C. This modification resulted in the exclusion of seven potentially nonimpacted stations for which there was inadequate power to distinguish significant effects relative to reference conditions. The excluded stations were Stations DR-05, DR-08, EW-03, and NS-04 in Elliott Bay, and Stations EW-14, SD-02, and SR-07 in Everett Harbor. By applying the guidelines in Appendix C to all stations in the SEDQUAL database, impacted stations were those that exhibited statistically significant mortality ($P \leq 0.05$) relative to reference conditions and exceeded 25 percent mortality (see rationale and recommendations in Appendix C for use of this criterion as a minimum level of concern). Twenty stations exhibited statistically significant mortality ($P \leq 0.05$) but were considered nonimpacted because mean mortality was ≤ 25 percent.

2.4.2 Benthic Infauna Data

Several modifications were also made for treatment of benthic infauna data to improve consistency of the results among the various pooled studies. As with the amphipod bioassay data, a comparisonwise significance level of $P \leq 0.05$ was used for all pairwise benthic comparisons. Power analyses were not conducted for benthic infauna because they were beyond the scope of this project. In lieu of a power analysis, and analogous to the amphipod bioassay guideline, a guideline was developed to ensure that significant benthic effects were of sufficient magnitude to be of concern as adverse impacts and to be discriminated statistically in most cases. Thus, only significant

effects ($P \leq 0.05$) that also exceeded a 50-percent reduction in major taxa abundance were considered impacts. Five stations (Station AP-01 from the Alki Point survey, Stations CI-20, RS-14, and SI-15 from the Commencement Bay survey, and Station WP-16 from the Metro TPPS survey; see Appendix B) exhibited significant benthic effects ($P \leq 0.05$) but were considered nonimpacted because depressions in abundance were ≤ 50 percent.

This guideline was derived partly from consideration of the natural variability of benthic infauna in relatively undisturbed environments of Puget Sound. Based on a summary of data from Lie (1968), Nichols (1975), and Word et al. (1984) in Tetra Tech (1987), the abundances of selected major taxa (Polychaeta, Mollusca, Crustacea) and total infauna may vary seasonally by roughly a factor of two (i.e., lowest mean abundances are roughly 50 percent of the highest mean abundances). In most cases, ≥ 50 percent reductions in mean abundance can be detected statistically ($P \leq 0.05$), whereas ≤ 30 percent reductions cannot be detected ($P > 0.05$). Finally, the guideline of 50 percent reduction in benthic infauna abundances provides a recommended level of environmental protectiveness for regulatory application with a reasonable balance between underprotection due to tolerance of major effects and overprotectiveness due to misclassification of nonimpacted sites as impacted. Nevertheless, a criterion based on 50 percent cannot be considered as sensitive or protective as a criterion based on a lower value (e.g., 30 percent depression).

2.4.3 Chemical Data

All detected chemical data entered in SEDQUAL after quality assurance review were included in AET calculations. These calculations were based on biological effect stations that passed the biological screening criteria summarized in the previous section.

A guideline was adopted for this report to address the concern over anomalously high chemical concentrations at nonimpacted stations (e.g., because of unusual matrix effects or low bioavailability). From a technical perspective (i.e., representativeness of data used in AET generation), the AET database was screened for biologically nonimpacted sediments exhibiting chemical concentrations that were anomalously high by a factor of 3 from that at the nonimpacted station exhibiting the next highest concentration (see Appendix C). From a management perspective, this guideline generates more protective (sensitive) sediment quality standards that may also be less efficient in only identifying problem sediments.

The purpose of this guideline was to acknowledge potentially nonrepresentative data for exceptional chemical matrices (e.g., slag, coal), or unusual biological conditions (e.g., extremely tolerant species under localized conditions). Only a limited number of stations were identified using this guideline. For the amphipod bioassay, this guideline affected 8 of 295 stations that passed biological guidelines (Section 2.4.1) and resulted in changes to the AET for nine chemicals where the ratio of the anomalous station to the nonimpacted station with the next highest concentration ranged from 3.2 to 14 (Table 6).

For benthic infauna, this procedure affected 4 of 205 stations that passed biological guidelines (Section 2.4.2) and resulted in changes to the AET for seven chemicals [including high molecular weight PAH (HPAH) as a class) where the ratio of the anomalous station to the nonimpacted station with the next highest concentration ranged from 3.0 to 20 (Table 6). Data currently excluded as anomalous are intended to be restored if and when confirming data become available.

TABLE 6. STATIONS ANOMALOUS WITH RESPECT TO SEDIMENT CHEMISTRY[a]

Survey	Station	Chemical	Ratio of Anomalous Station to Next Highest Station
Chemical Anomalies for the Amphipod Mortality Bioassay			
CBMSQS	CI-16	N-nitrosodiphenylamine	4.5
	HY-12	Di-n-butyl phthalate	3.2
EBCHEM	DR-10	DDE	4.2
	DR-12	Arsenic	4.8
	NS-06	Di-n-octyl phthalate	14.3
	SS-04	Dibenzoanthracene	3.6
	SS-10	Chromium	4.0
	WW-14	Antimony	7.1
		Lead	14.3
Chemical Anomalies for Benthic Infaunal Abundance			
EBCHEM	AB-01	Mercury	14.3
	SS-10	Chromium	4.2
TPPS3AB	WP-11 (3/15/82)	Benzo(a)pyrene	4.8
		Benzo(ghi)perylene	3.4
		HPAH	3.0
		Phenanthrene	3.4
		Total benzofluoranthenes	7.1
	WP-14 (7/15/82)	Di-n-octyl phthalate	20.0

[a] Nonimpacted stations that potentially set AET but exceed the next highest nonimpacted station for one or more chemicals by greater than a factor of three (note: Benthic effects at EBCHEM Station SS-10 were statistically significant at $P \leq 0.05$ but the observed depressions were <50 percent, so the station was classified as nonimpacted).

3.0 RESULTS

3.1 PREDICTIVE RELIABILITY OF 1986 PUGET SOUND AET

1986 AET developed for PSEP/PSDDA (Beller et al. 1986; see Table 1) were used to independently evaluate new data from Ecology's Preliminary Investigation of Eagle Harbor, and PSEP sampling programs in Elliott Bay and Everett Harbor. Results of these comparisons are summarized in Figure 4 and Table 7. The purpose of this comparison was to further test the predictive sensitivity and efficiency of the 1986 Puget Sound AET in identifying problem sediments. Only amphipod bioassay and benthic infaunal abundance data were collected in these new surveys.

The sensitivity of 1986 benthic infauna AET in correctly identifying impacted benthic infauna stations ranged from 71 to 100 percent for the three surveys considered independently, and summed to 82 percent (58/71 impacted stations) for the three surveys combined. This sensitivity is comparable to that reported overall for the surveys used in 1986 to calculate these AET (82 percent; Beller et al. 1986). The predictive efficiency of the 1986 AET applied to the new surveys ranged from 50 to 100 percent and summed to 74 percent (58/78 predicted impacts were actual impacts). Overall, the greatest reliability was found in Everett Harbor (especially the contaminated East Waterway, where 100 percent sensitivity and 100 percent efficiency were achieved). Selected areas of Elliott Bay (Duwamish River East Waterway and the Seattle South Waterfront) exhibited lower efficiency (50-60 percent) but high sensitivity (100 percent).

The sensitivity of 1986 amphipod bioassay AET in correctly identifying impacted amphipod bioassay stations generally ranged from 63 to 100 percent, and totaled 66 percent (35/53 impacted stations) for the three surveys combined. This sensitivity is also comparable to an overall sensitivity of 54 percent for the original surveys used to calculate these AET. A low sensitivity of 33 percent was observed for samples from the upper Duwamish River in the PSEP Elliott Bay survey. However, the corresponding efficiency for the upper Duwamish River stations was 100 percent (i.e., the two predictions of impacts were both correct). The efficiency for other areas was variable, ranging from 20 percent (Eagle Harbor) to 83 percent, and summed to 56 percent (35/63 predicted impacts were actual impacts). Overall, the greatest predictive reliability was again found in the East Waterway of Everett Harbor (100 percent sensitivity and 80 percent efficiency).

3.2 RELIABILITY OF 1988 PUGET SOUND AET

The sensitivity of 1988 amphipod bioassay and benthic infauna AET (see Table 2; Section 2.2) in identifying impacted stations (for the appropriate biological indicator) from all available Puget Sound surveys is evaluated in this section. The sensitivity results are compared with the overall sensitivity of 1986 Puget Sound AET discussed in the previous section. The purpose of this comparison is to evaluate the change in sensitivity after approximately doubling the size of the database used to derive amphipod bioassay and benthic infauna AET by incorporation of data from the major urban bays programs sponsored by PSEP. Because chemical data for all nonimpacted stations in the database were used in the generation of these 1988 AET, they are by definition, 100 percent efficient in only predicting impacted stations as impacted (see Sections 2.1

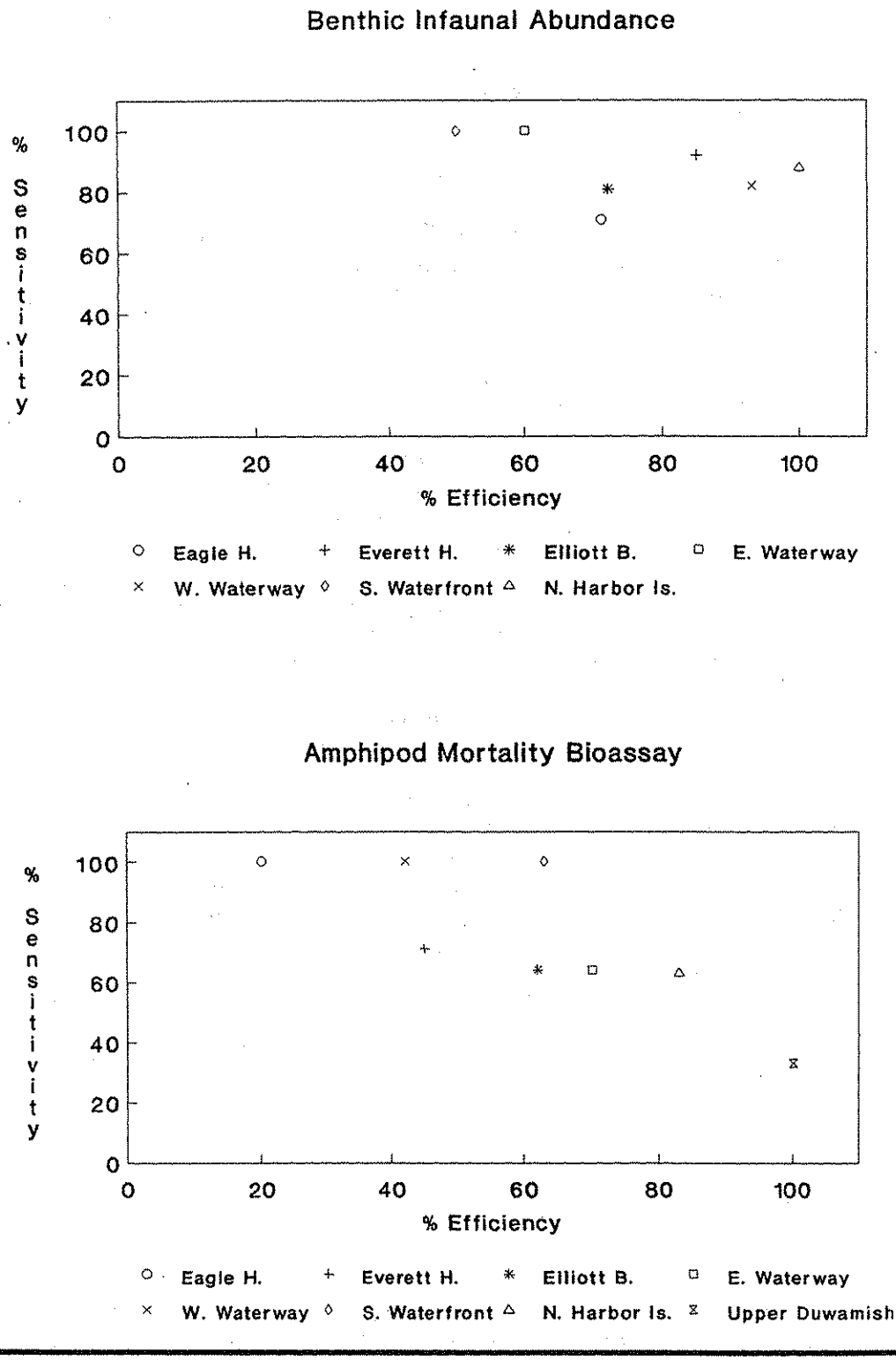


Figure 4. Reliability of 1986 Puget Sound AET in Eagle Harbor, Everett Harbor, and Elliott Bay, and selected geographic areas of Elliott Bay.

TABLE 7. RELIABILITY OF 1986 PUGET SOUND AET IN PREDICTING ADVERSE BIOLOGICAL EFFECTS FOR NEW SURVEYS

Location	Survey	Station Code	Sensitivity	Efficiency
Benthic AET vs. Benthic Infaunal Abundance Stations				
Eagle Harbor	EHCHEM	All	71% (5/7)	71% (5/7)
Elliott Bay	EBCHEM	All	81% (42/52)	72% (42/58)
Duwamish E. Waterway	EBCHEM	EW	100% (9/9)	60% (9/14)
Duwamish W. Waterway	EBCHEM	WW	82% (14/17)	93% (14/15)
S. Waterfront	EBCHEM	SS	100% (5/5)	50% (5/10)
N. Harbor Island	EBCHEM	NH	88% (7/8)	100% (7/7)
Everett Harbor	EVCHEM	All	92% (11/12)	85% (11/13)
Everett E. Waterway	EVCHEM	EW	100% (6/6)	100% (6/6)
Amphipod AET vs. Amphipod Mortality Bioassay Stations				
Eagle Harbor	EHCHEM	All	100% (1/1)	20% (1/5)
Elliott Bay	EBCHEM	All	64% (29/45)	62% (29/47)
Duwamish E. Waterway	EBCHEM	EW	64% (7/11)	70% (7/10)
Duwamish W. Waterway	EBCHEM	WW	100% (5/5)	42% (5/12)
S. Waterfront	EBCHEM	SS	100% (5/5)	63% (5/8)
N. Harbor Island	EBCHEM	NH	63% (5/8)	83% (5/6)
Upper Duwamish R.	EBCHEM	DR	33% (2/6)	100% (2/2)
Everett Harbor	EVCHEM	All	71% (5/7)	45% (5/11)
Everett E. Waterway	EVCHEM	EW	100% (4/4)	80% (4/5)

and 2.2). This constraint on 100 percent efficiency does not ensure high sensitivity, which requires correct prediction of impacted stations.

Results for the sensitivity of the 1988 AET are presented in Figure 5 and Table 8 for different geographic areas and surveys (NOTE: 4 of the 7 areas shown are major geographic areas located within Elliott Bay; the average sensitivity for all of Elliott Bay is also shown). The sensitivity of the 1988 benthic infauna AET in correctly identifying impacted stations generally ranged from 55 to 100 percent, and totaled 75 percent for all surveys combined. The highest sensitivity was found in the Everett Harbor and Commencement Bay surveys (93 to 100 percent). The single impacted station in the Alki Point survey was not predicted, resulting in 0 percent sensitivity for this one area.

The sensitivity of the 1988 amphipod AET ranged from 71 to 100 percent for the major urban bay and Superfund surveys in Commencement Bay, Eagle Harbor, Elliott Bay, and Everett Harbor. Comparisons to stations from the 1985 EIGHTBAY survey resulted in a notably low sensitivity of 17 percent and the single impacted subsurface sediment sample from the EVERETT1 dredging survey was also not predicted. An intermediate level of sensitivity (43 to 50 percent) was observed for the DUWRIV1 and DUWRIV2 dredging surveys. Potential reasons for these differences are discussed in later sections; however, the influence of physical factors (e.g., grain size) on bioassay results and the availability of low detection limits (i.e., <50 ppb) for a wide range of organic chemicals appear to be more important than geographic location.

A summary of sensitivity and efficiency results for several reliability tests of AET are shown in Table 9. The table includes data for AET normalized to dry weight and to TOC (see Section 3.2.2 for a comparison of these two sets of results). The overall sensitivity of 1988 amphipod bioassay AET normalized to dry weight is slightly higher than that for analogous 1986 AET (two additional impacted stations are now predicted; Table 9), although AET have increased for many chemicals (e.g., PAH). This net increase in sensitivity rather than a decrease is explained by the following four factors:

1. Increasing an AET value does not necessarily mean that previously predicted stations will no longer be predicted, because the concentration of chemicals at such stations may still exceed the updated AET value. Alternatively, the impacted stations may still be predicted by AET for other chemicals that are also detected at elevated concentrations at the station.

2. Implementation of consistent rules for treatment of anomalous biological and chemical data (Section 2.4 and Appendix C) in developing the 1988 AET resulted in the lowering of AET values relative to corresponding 1986 AET for five chemicals used in predictions (see footnote h in Table 2).

3. 1988 AET are available for some chemicals that either were not used in predictions by 1986 AET (e.g., antimony, chromium; see Section 2.3 and Appendix F) or were not analyzed for in the surveys used to establish 1986 AET (e.g., new data are available for resin acids and chlorinated phenols in the Everett Harbor survey and selected tentatively identified compounds detected in the Everett Harbor and Elliott Bay surveys).

4. Some 1986 AET are "greater than" values (i.e., preliminary AET; see Section 2.2) because no impacted stations exceeded the highest concentration at a nonimpacted station in the surveys used to establish 1986 AET. These preliminary AET are not used in determining sensitivity. After

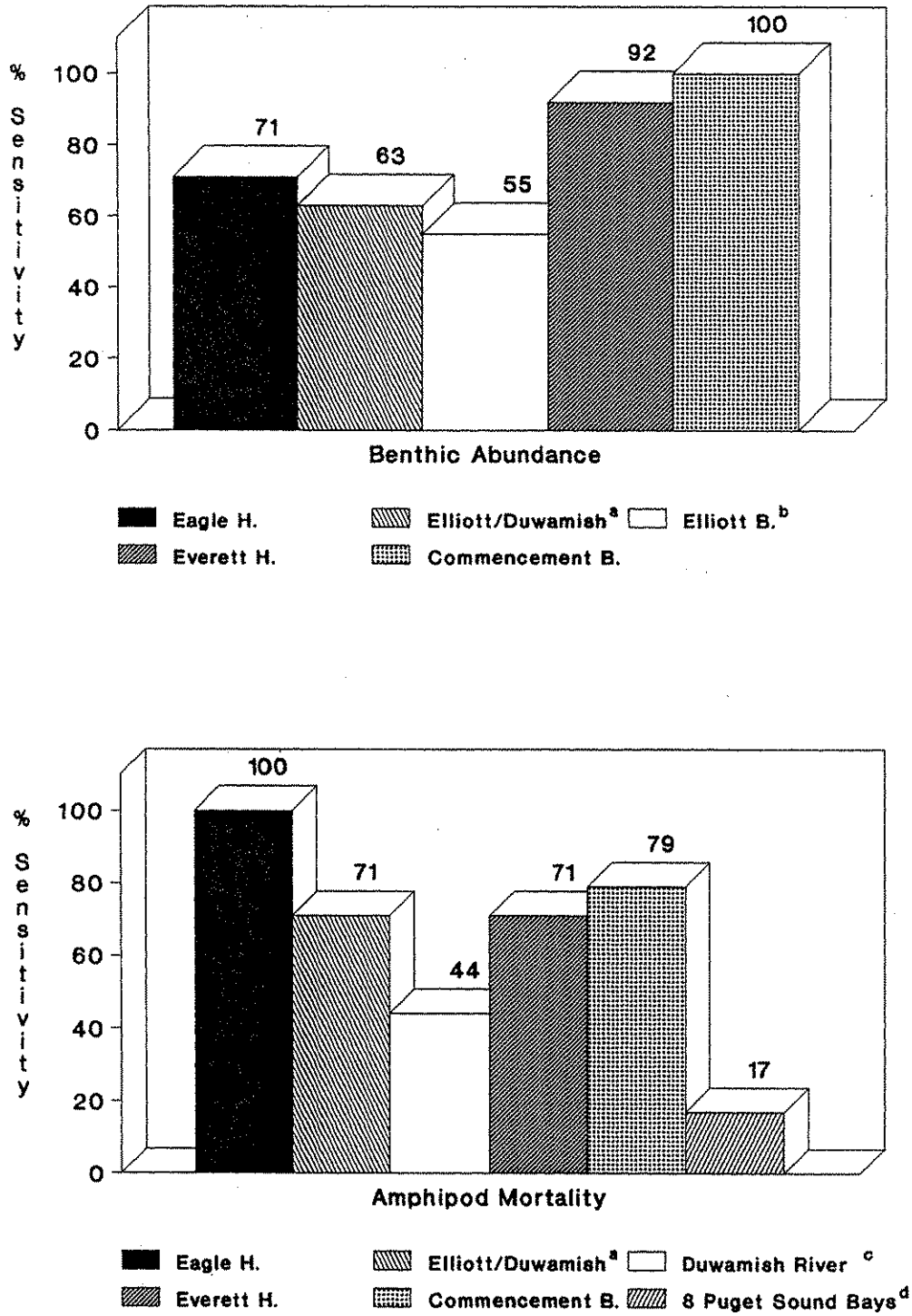


Figure 5. Sensitivity of 1988 Puget Sound amphipod and benthic infaunal AET by geographic area.

TABLE 7. RELIABILITY OF 1986 PUGET SOUND AET IN PREDICTING ADVERSE BIOLOGICAL EFFECTS FOR NEW SURVEYS

Location	Survey	Station Code	Sensitivity	Efficiency
Benthic AET vs. Benthic Infaunal Abundance Stations				
Eagle Harbor	EHCHEM	All	71% (5/7)	71% (5/7)
Elliott Bay	EBCHEM	All	81% (42/52)	72% (42/58)
Duwamish E. Waterway	EBCHEM	EW	100% (9/9)	60% (9/14)
Duwamish W. Waterway	EBCHEM	WW	82% (14/17)	93% (14/15)
S. Waterfront	EBCHEM	SS	100% (5/5)	50% (5/10)
N. Harbor Island	EBCHEM	NH	88% (7/8)	100% (7/7)
Everett Harbor	EVCHEM	All	92% (11/12)	85% (11/13)
Everett E. Waterway	EVCHEM	EW	100% (6/6)	100% (6/6)
Amphipod AET vs. Amphipod Mortality Bioassay Stations				
Eagle Harbor	EHCHEM	All	100% (1/1)	20% (1/5)
Elliott Bay	EBCHEM	All	64% (29/45)	62% (29/47)
Duwamish E. Waterway	EBCHEM	EW	64% (7/11)	70% (7/10)
Duwamish W. Waterway	EBCHEM	WW	100% (5/5)	42% (5/12)
S. Waterfront	EBCHEM	SS	100% (5/5)	63% (5/8)
N. Harbor Island	EBCHEM	NH	63% (5/8)	83% (5/6)
Upper Duwamish R.	EBCHEM	DR	33% (2/6)	100% (2/2)
Everett Harbor	EVCHEM	All	71% (5/7)	45% (5/11)
Everett E. Waterway	EVCHEM	EW	100% (4/4)	80% (4/5)

TABLE 9. SENSITIVITY AND EFFICIENCY RESULTS FOR 1986 AND 1988 PUGET SOUND AET AND INDEPENDENT AET (SINGLE STATION DELETION TEST)

[normalized either to dry weight (DW) or total organic carbon (TOC)]

AET Dataset[a]	Number of Stations Evaluated	Sensitivity	Efficiency	Overall Reliability
Benthic Infaunal Abundance				
1986 Puget Sound AET (DW)	201	76% (82/108)	82% (82/100)	78% (157/201)
1988 Puget Sound AET (DW)	201	75% (81/108)	100% (81/81)[b]	87% (174/201)
1988 AET (mixed DW/TOC)	201	77% (83/108)	100% (83/83)[b]	88% (176/201)
1988 AET (TOC only)	201	76% (82/108)	100% (82/82)[b]	87% (175/201)
Independent AET[c] (DW)	201	75% (81/108)	72% (81/112)	71% (143/201)
Amphipod Mortality Bioassay				
1986 Puget Sound AET (DW)	287	56% (59/106)	69% (59/86)	74% (213/287)
1988 Puget Sound AET (DW)	287	58% (62/106)	100% (62/62)[b]	85% (243/287)
1988 AET (mixed DW/TOC)	287	55% (58/106)	100% (58/58)[b]	83% (239/287)
1988 AET (TOC only)	287	45% (48/106)	100% (48/48)[b]	80% (229/287)
Independent AET[c] (DW)	287	57% (60/106)	67% (60/90)	74% (211/287)
Microtox Bioassay				
1986 Puget Sound AET[d] (DW)	50	93% (27/29)	100% (27/27)[b]	96% (48/50)
1986 AET (mixed DW/TOC)	50	93% (27/29)	100% (27/27)[b]	96% (48/50)
1986 AET (TOC only)	50	83% (24/29)	100% (24/24)[b]	90% (45/50)
Independent AET[c] (DW)	50	93% (27/29)	61% (27/44)	62% (31/50)
Oyster Larvae Abnormality				
1986 Puget Sound AET[d] (DW)	56	88% (15/17)	100% (15/15)[b]	96% (54/56)
1986 AET[d] (mixed DW/TOC)	56	88% (15/17)	100% (15/15)[b]	96% (54/56)
1986 AET (TOC only)	56	71% (12/17)	100% (12/12)[b]	91% (51/56)
Independent AET[c] (DW)	56	88% (15/17)	37% (15/41)	50% (28/56)

[a] Results for TOC-normalized AET are included in this table for comparison to the results for dry weight-normalized AET. The mixed DW/TOC dataset includes TOC-normalized AET generated using 1988 data for nonionic organic compounds and dry weight-normalized AET generated for the remaining chemicals (metals and ionizable organic compounds). The TOC only dataset includes TOC-normalized AET generated using 1988 data for all chemicals (i.e., metals and nonionic and ionizable organic compounds).

TABLE 9. (Continued)

[b] By definition, efficiency is 100 percent because all Puget Sound stations for each indicator were included in the calculation of these AET.

[c] Cumulative results for (1) deleting a station from the AET database; (2) recalculating AET using remaining stations; (3) predicting effects at deleted station; (4) restoring the deleted station to the AET database; (5) repeating for each station in the database (see text for discussion of why this test affects efficiency more than sensitivity).

[d] No new data were available to update 1986 AET for oyster larvae and Microtox bioassays; the results shown for dry weight- and TOC-normalized AET for these indicators differ somewhat from those previously reported by Beller et al. (1986) because more chemicals were included in the analysis (e.g., including tentatively identified organic compounds and conventional sediment variables such as TOC, grain size, sulfides).

addition of new survey data, 1988 AET for some of these chemicals (e.g., pentachlorophenol in Tables 1 and 2) became definitive AET because at least one impacted station in the new surveys exceeded the highest concentration at a nonimpacted station. Such AET are used in determining sensitivity.

3.2.1 Estimation of Predictive Efficiency and Sensitivity

Because the efficiency of the 1988 AET is 100 percent by definition for the entire Puget Sound database used in their calculation, predictive efficiency was estimated by the following procedure. For each biological indicator, sequentially delete a single station from the total database, recalculate AET for the remaining data set, and predict biological effects for the single deleted station. The predictive efficiency is then the cumulative results for the sequential deletions of single stations. For example, the 287-sample database for amphipod bioassay results can be used to provide a 286-sample independent database for predicting (in sequence) effects on all 287 samples.

Reliability results for these single deletion tests are compared with those for 1986 and 1988 AET for amphipod bioassay and benthic infaunal abundance in Figure 6 and Table 9. Similar results for oyster larvae and Microtox bioassays are also presented for comparison in Table 9.

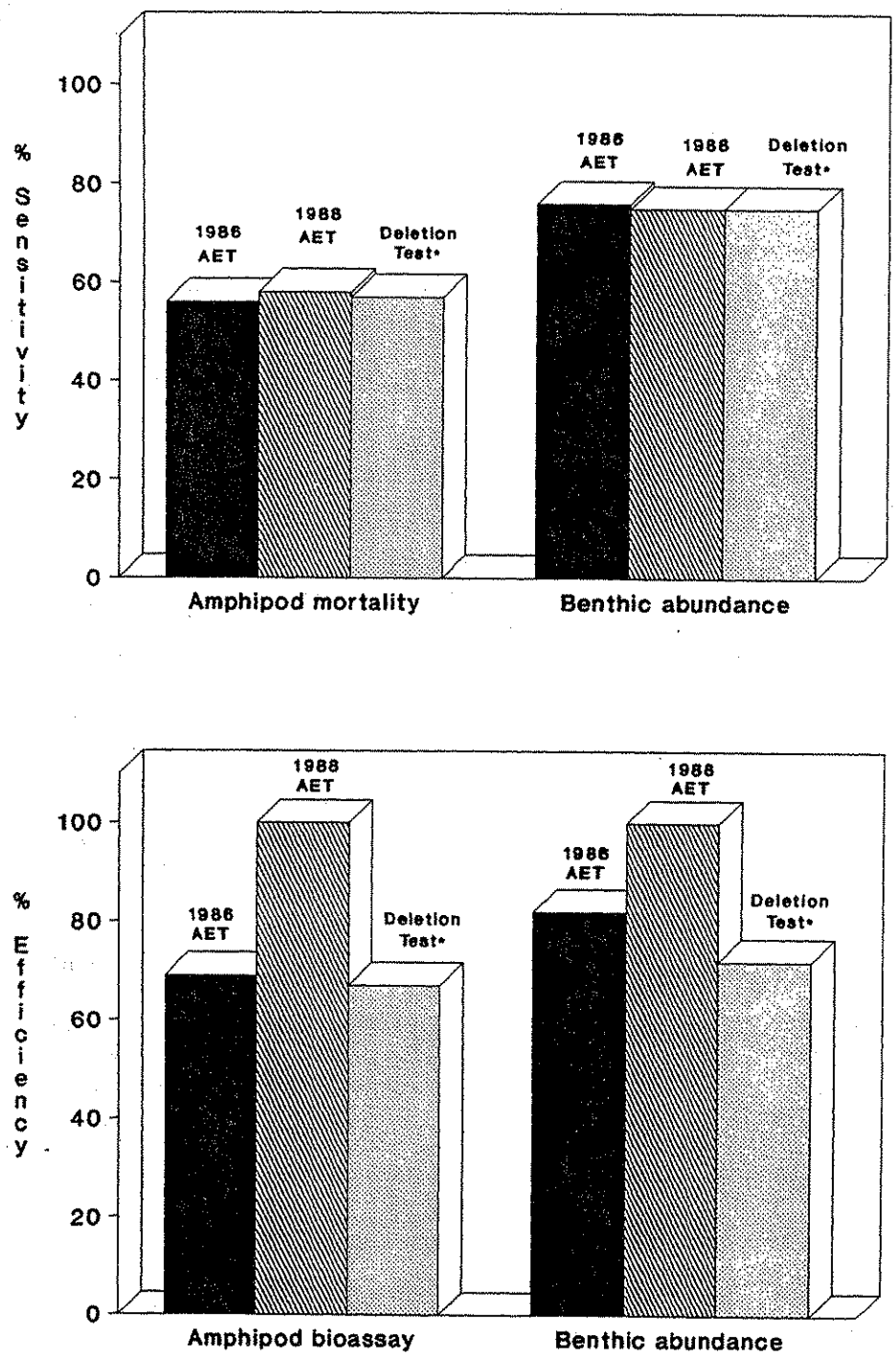
A consideration of the calculation of sensitivity (the percent of impacted stations that are correctly predicted) demonstrates why there is little or no decrease in predictive sensitivity based on sequential deletions of single stations (Figure 6):

- For a given data set, single deletions of nonimpacted stations do not alter sensitivity because sensitivity only measures the ability to predict impacted stations.

- Sensitivity can only change if there is a change in AET, which are established by nonimpacted stations. Single deletions of impacted stations therefore do not change AET. The only exception is the case that a deleted station is the only impacted station to exceed an AET, in which case the AET is no longer considered "definitive" when that impacted station is deleted. This case does not affect sensitivity for benthic infauna AET because the few such stations are still predicted by at least one other benthic infauna AET exceeded by multiple impacted stations. The sensitivity of amphipod AET is reduced by less than 2 percent because two amphipod AET (silver and 2,3,5-trimethylnaphthalene) are affected by this exception.

In practice, this complete enumeration procedure is too time-consuming for routinely processing large data sets using available database programs. A calculating algorithm was used that yields the same results by accounting for the following rules concerning efficiency (the percent of correct predictions out of all predictions of impacts):

- For a given data set, efficiency changes only as the result of deleting nonimpacted stations that uniquely establish an AET value. When such stations are deleted and AET are recalculated, the resulting lower AET incorrectly predicts the deleted station as an impacted station, thus lowering predictive efficiency from 100 percent.



*Deletion tests represent results for independent prediction of effects on all stations in database (see text).

Figure 6. Reliability of 1986 Puget Sound AET, 1988 AET, and single station deletion test.

- For all other stations in the data set, no incorrect predictions are made when they are deleted individually, so efficiency does not change.

The resulting efficiency summarized in Figure 6 is only an approximation of predictive efficiency and is expected to be an underestimate when a large database representing a wide range of contaminant conditions has been assembled (as for Puget Sound). This is true because the estimating method essentially assumes that none of the nonimpacted stations currently establishing AET would be represented in the predictive database, which is unlikely.

Additional checks on the predictive reliability of 1988 AET that were performed for PSDDA are provided in Appendix G. These results are based on comparisons to amphipod bioassay stations from independent dredging projects compiled by PSDDA but because of data limitations described in Appendix G, these data have not been recommended for inclusion in the sediment quality values database.

3.2.2 Relative Performance of Dry Weight- and TOC-Normalized AET

An interesting result of validation tests with Puget Sound data is the roughly equivalent predictive success of dry weight-normalized AET and organic carbon-normalized AET (Table 9 and Figure 7). This finding is unexpected in light of existing equilibrium partitioning theory and associated laboratory bioassay and sorption studies, which strongly support the use of TOC normalization for nonpolar organic chemicals. Possible explanations for the relative performance of dry weight- and organic carbon-normalized AET are presented in this section.

Dry weight-normalization simply assumes that the overall burden of a contaminant in sediment is a predominant factor influencing toxicity to exposed organisms (although organic carbon content may be a secondary factor). Because dry weight-normalization does not focus on any specific solid fraction of the sediment (e.g., organic matter, fine particles), it essentially averages among sediment fractions in terms of sorptive affinity and relationship to bioavailability. If specific sediment fractions consistently mediate toxicity in most samples, then averaging across all sediment fractions potentially reduces the representativeness of dry weight-normalized AET. Alternatively, if specific sediment fractions do not consistently mediate toxicity in most samples, normalization to dry weight may better account for the variability of contaminant-toxicity relationships among environmental samples than normalization to these sediment fractions.

Organic carbon normalization assumes that organic matter in sediments is a generic sink for nonionic organic contaminants and is a predominant factor influencing the bioavailability and toxicity of these compounds to exposed organisms. Laboratory bioassay and sorption studies provide a mechanistic rationale for organic carbon normalization. Simply stated, bioavailability and toxicity of nonionic organic chemicals in sediments appear to correspond to interstitial water concentrations, and, under equilibrium conditions, the distribution of nonionic organic compounds between sedimentary organic matter (represented by organic carbon content) and interstitial water should be constant (and can be expressed as K_{oc}). For sediments with a given bulk concentration of a nonionic organic chemical, increases in organic carbon content should correspond to proportional decreases in interstitial water concentrations of that chemical. Hence, as sediment organic carbon content increases, toxicity "threshold" values expressed per gram of bulk sediment should decrease. If contaminant concentrations are normalized to organic carbon content, threshold values should be constant for that contaminant in all sediments regardless of organic carbon content. Thus, use of TOC-normalized data

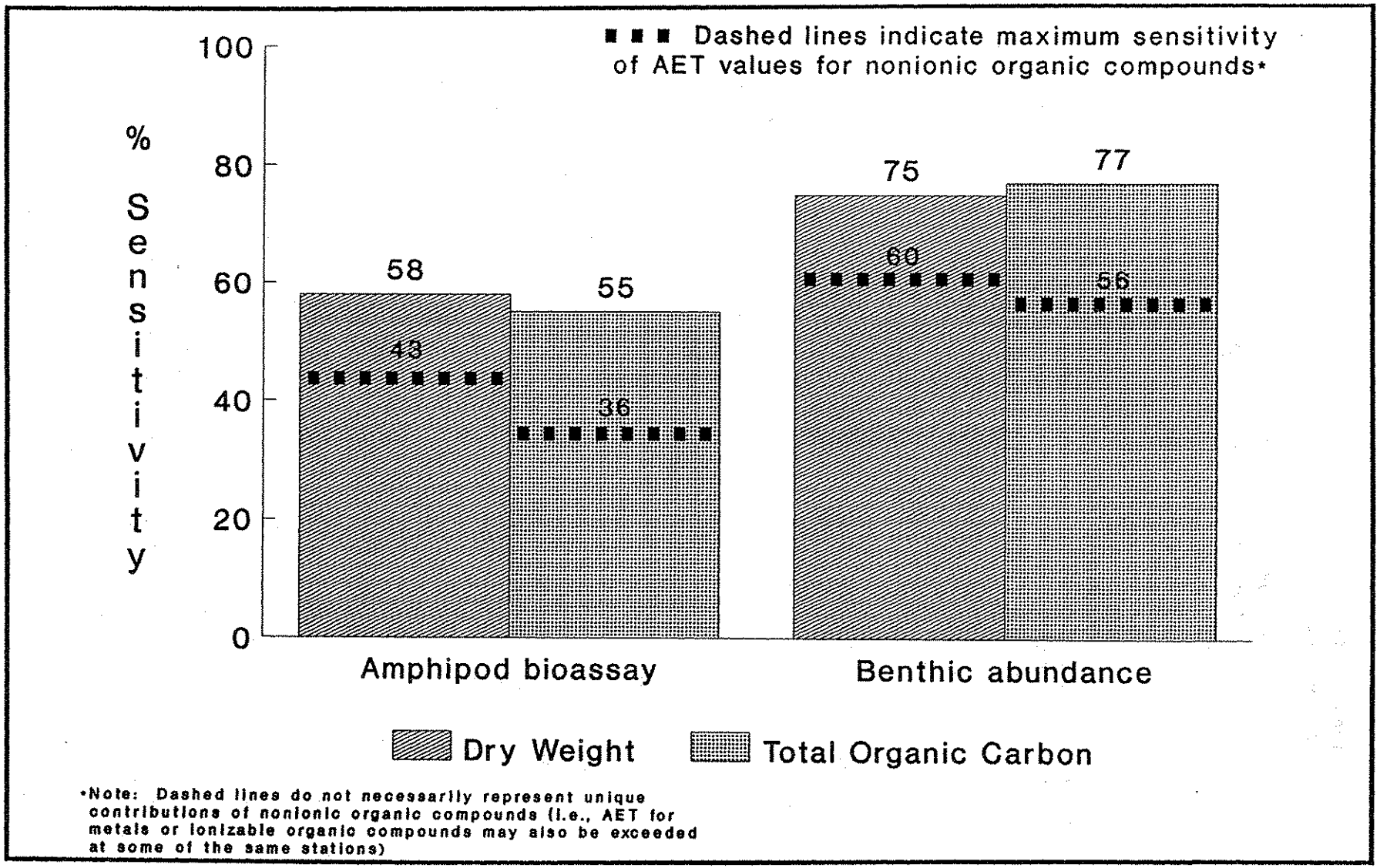


Figure 7. Sensitivity of dry weight- and total organic carbon-normalized AET. [Efficiency is 100 percent by constraint for both tests.]

could be expected to result in a more consistent estimation of sediment toxicity thresholds than use of dry weight-normalized data for generation of AET for nonionic organic compounds.

A study by Adams et al. (1985) illustrates the rationale for organic carbon normalization. Adams et al. (1985) conducted a series of bioassays in which freshwater midges *Chironomus tentans)* were exposed to water, sediments (with various levels of organic carbon content), and food contaminated with Kepone. No-effect concentrations based on total sediment Kepone concentrations increased in proportion to total organic carbon content of sediments, whereas no-effect levels based on interstitial water Kepone concentrations were relatively constant regardless of sediment concentration. The authors suggested that no-effect concentrations should be based on interstitial water concentrations and sediment organic carbon content, not on bulk sediment weight.

The AET approach is empirical and does not favor one mechanistic explanation over any other but can operate whether one or a combination of assumptions is appropriate. The interpretation of operative mechanisms is, for the AET approach, an *a posteriori* inference based on the results of validation tests. For contaminated sediments in the environment, organic carbon normalization might not offer advantages in predictive success over dry weight normalization if the following hypotheses are applicable to environmental sediment samples:

- Sediment/interstitial water systems are not typically at equilibrium in the environment as a result of impeded contaminant exchange
- Sediment organic matter occurring in the environment does not have uniform affinity for nonionic organic compounds.

These hypotheses are discussed further in the following sections.

Sediment-Water Contaminant Exchange--Based on the mechanistic rationale for organic carbon normalization, an assumed advantage of TOC-normalized AET is the constant relationship between sedimentary organic matter, interstitial water concentration, and sediment toxicity that should exist under equilibrium conditions. Based in part on evidence from laboratory studies, it is plausible that equilibrium could be difficult to attain in the environment because of kinetic aspects of sorption/desorption processes. The attainment of equilibrium requires a relatively rapid transfer of a contaminant between various phases in a system. Studies of sorption/desorption have demonstrated that attainment of equilibrium of a nonionic organic compound between sediment and aqueous phases can take weeks, months, or longer (e.g., Karickhoff 1984, Karickhoff and Morris 1985b). The existence of reversible (rapidly exchangeable) and irreversible (highly retarded) components of contaminant loadings in sediments has been postulated by several investigators based on hysteresis in sorption/desorption isotherms or direct observation of desorption kinetics (e.g., DiToro and Horzempa 1982; Karickhoff and Morris 1985b). In laboratory-spiked sediments studied by Karickhoff and Morris (1985b), one half or more of the total sorbed contaminant concentration was irreversible. For highly hydrophobic compounds and systems with high solids concentrations, the irreversible fraction increased to $\geq$90 percent in some cases (Karickhoff and Morris 1985b). For the irreversible component, the attainment of equilibrium distributions could be very slow, on the order of years for highly hydrophobic compounds (i.e., relatively widespread contaminants such as PAH with five or more rings and PCB congeners with six or more chlorine atoms) (Karickhoff and Morris 1985b).

Very few studies have been conducted with field-collected sediments to examine the kinetic aspects of contaminant exchange between sediment and ambient water. However, in diffusion studies with field-contaminated sediments, Fisher et al. (1983) reported apparent diffusion coefficients for trichloro-, tetrachloro-, and pentachlorobiphenyls that were one to three orders of magnitude lower than observed for laboratory-spiked hexachlorobiphenyl (DiToro et al. 1985); the laboratory-spiked PCB congener appeared to exhibit fully reversible behavior. DiToro et al. (1985) proposed that the low apparent diffusion of PCB congeners observed by Fisher et al. (1983) could be explained by the fact that 90 to 99.9 percent of the sediment loading of PCBs was irreversibly sorbed in those field-collected sediments. Such a proportion of irreversibly sorbed contaminants in the environment could result in considerable deviations from equilibrium conditions. Relatively large and variable deviations from equilibrium in environmental samples would result in considerable variability in the relationship between organic carbon content and bioavailability or toxicity on a sample-by-sample basis, reducing the potential advantage of TOC normalization.

Physical explanations for retarded sediment/water exchange of contaminants and deviations from equilibrium include entrainment or trapping of contaminants in refractory matrices such as fecal pellets (Karickhoff and Morris 1985a) and humic matter (Freeman and Cheung 1981). In a study of sediments collected in Puget Sound, Prahl and Carpenter (1983) observed that PAH were disproportionately concentrated in certain fractions of refractory sedimentary organic matter (e.g., charcoal fragments and vascular plant detritus, such as lignin). This disproportionality suggests that PAH may not have been at equilibrium within the sediment phase or, alternatively, that different kinds of organic matter may have different affinities for PAH.

The Uniformity of Organic Matter--For a given nonionic organic chemical, consistent partitioning between organic matter and interstitial water in sediments from different areas requires that organic matter have a consistent affinity for the chemical. A number of studies indicate that organic matter can be considered uniform to a first approximation, based on the derivation of single K_{oc} values for ranges of sediments and soils (e.g., Karickhoff 1984 and references therein). However, direct studies of the associations of nonionic organic compounds with dissolved (Gauthier et al. 1987; Carter and Suffet 1985) and particulate (Diachenko 1981) humic materials indicate that affinities can vary considerably as a function of the source and properties of the organic matter. For example, K_{oc} values for pyrene in dissolved humic materials from various sources differed by as much as an order of magnitude (Gauthier et al. 1987). The aromatic vs. aliphatic character of organic matter has been related to binding affinity (Diachenko 1981; Gauthier et al. 1987) and is likely an important structural difference between humic materials derived from terrestrial vs. marine sources (e.g., Hatcher et al. 1980, 1981) in estuarine sediments. Variability in organic matter alone may not account for the roughly equivalent performance of dry weight- and organic carbon-normalized AET, as variability within the sedimentary pool of organic matter does not necessarily favor normalization to bulk sediment rather than bulk organic carbon.

3.3 EVALUATION OF FACTORS AFFECTING AET RELIABILITY

As part of the refinement of sediment quality values, a number of tests were performed to evaluate the reliability of AET as follows:

- Identification of characteristics of stations that do not conform to predictions of AET (i.e., nonimpacted stations that are predicted to be impacted by AET, or impacted stations that are not predicted by AET)

- Evaluation of the potential influence of geographic location on the reliability of AET predictions in Puget Sound
- Estimation of the minimum numbers of stations needed to establish empirically reliable AET
- Identification of subsets of chemical AET that account for most of the reliability of the Puget Sound AET.

Results for each of these evaluations are summarized in Sections 3.3.1-3.3.4. Based on these results, recommendations for development and use of AET are made in Section 4.0 (SUMMARY). Unless otherwise specified, the discussion of AET reliability is based on comparisons for 1988 AET for amphipod bioassay and benthic infaunal abundance indicators because new data are not available for the other indicators.

3.3.1 Characteristics of Biologically Impacted Stations Not Predicted as Impacts by AET

The purpose of this section is to identify the characteristics of biologically impacted stations for which no identified chemical exceeds the AET for that biological indicator. Characteristics of stations that do not exhibit biological effects but which are predicted to be impacted according to the AET approach have been evaluated based on interim sediment quality values developed in 1986 for PSDDA (see Volume II of this Sediment Quality Values Refinement series; Barrick et al. 1988).

In the 334-station database, biological effects were found at 49 amphipod bioassay or benthic infauna stations at which no chemical concentration exceeded a 1988 AET. These incorrect predictions included 29 of 287 stations for the amphipod mortality bioassay (Table 10) and 24 of 201 stations for effects on benthic macroinvertebrate communities (Table 11). Both kinds of effects were found at five of the impacted stations that were not predicted by AET. Chemical and biological data for each of these stations were reviewed to assess possible reasons for the incorrect predictions including the following:

- The indicated adverse biological effect resulted from factors other than toxic chemical exposures (e.g., physical disruption, grain size distribution)
- Unidentified chemicals accounted for the sediment toxicity
- High chemical detection limits precluded an assessment of whether the AET was exceeded
- The biological effect was incorrectly classified as significant because of a Type I statistical error.

Amphipod Bioassay Stations--DeWitt et al. (1988) have shown that elevated mortality can result in the amphipod mortality bioassay by exposing test organisms to sediments having a high percentage of fine-grained material. These responses were found to occur in the absence of apparent chemical contamination and were thought to result from the physical characteristics of the sediment or some other natural variable that correlated with the physical characteristics. In the SEDQUAL database, significant amphipod mortality was associated with 35 of 75 stations (47 percent) with a relatively

TABLE 10. CHARACTERISTICS OF STATIONS AT WHICH SIGNIFICANT AMPHIPOD MORTALITY WAS FOUND BUT NO EFFECTS WERE PREDICTED BY 1988 AMPHIPOD AET

Embayment	Survey	Station	Percent Mortality[a]	Key Distinguishing Factors[b,c]
Elliott Bay	EBCHEM	DR-14	32	% fines = 80.6
		EW-10	58	% fines = 80.3; DL
		KG-02	37	DL (6% fines)
		KG-03	27	DL (49% fines)
		KG-09[d]	33	DL (51% fines)
		KG-11	32	DL (8% fines)
		NH-02	45	DL (31% fines)
		NH-09	58	DL (11% fines)
		NH-11	55	no obvious factors (29% fines)
		NS-08[d]	82	% fines = 83.9; DL
		WW-08[d]	41	DL (58% fines)
Sinclair Inlet	EIGHTBAY	SC-08	38	% fines = 89.8; DL
		SC-17	26	% fines = 78.2; DL
		SC-18	32	% fines = 77.8; DL
Case Inlet		CS-01	28	% fines = 89.4; DL
		CS-11	28	DL (39% fines)
		CS-15	43	% fines = 81.3; DL
		CS-17	41	% fines = 78.1; DL
Dabob Bay		DB-07	26	DL (49% fines)
		DB-15	36	% fines = 89.7; DL
Samish Bay		SM-03	47	% fines = 80.9; DL
		SM-07	25	% fines = 84.9; DL
		SM-20	31	% fines = 87.2; DL
Everett Harbor		EV-02	29	% fines = 83.1; DL
		EV-11	26	no obvious factors (69% fines)
Bellingham Bay		BH-05	34	% fines = 96.6; DL
		BH-23	58	% fines = 95.3; DL
Everett Harbor	EVCHEM	NG-06	43	no obvious factors (7% fines)
Everett Harbor	EVERETT1	EV-24	40	few chemicals measured (63% fines)

[a] Percent mortality observed at each anomalous station.

[b] Percent fine-grained material (i.e., <63 um).

[c] DL = Detection limits of at least one chemical exceeded the 1988 amphipod bioassay AET for that chemical.

[d] Station predicted by 1986 amphipod AET but not 1988 amphipod AET

TABLE 11. CHARACTERISTICS OF STATIONS AT WHICH SIGNIFICANT BENTHIC EFFECTS WERE FOUND BUT NO EFFECTS WERE PREDICTED BY 1988 BENTHIC INFAUNA AET

Embayment	Survey	Station	Taxon[a]	Key Distinguishing Factors[b,c]
Elliott Bay	EBCHEM	AB-03	P	% coarse = 95.3; DL
		KG-07[d]	M	% coarse = 83.3; DL
		KG-08[d]	M	% coarse = 90.8; DL
		KG-11	M	% coarse = 92.5; DL
		MG-01	P	DL (25% coarse)
		MG-02	P	% coarse = 95.5; DL
		MG-03	P	% coarse = 77.5; DL
		MG-04	P	DL (14% coarse)
		NH-01[d]	M	% coarse = 81.1; DL
		NH-02[d]	M	DL (69% coarse)
		NH-11	P	% coarse = 71.3; DL
		NS-08[d,e]	T,M,C	DL; (16% coarse)
		WW-01	M	no obvious factors (60% coarse)
		WW-03	T,P,M,C	% coarse = 93.4
		WW-08[d]	M,C	no obvious factors (42% coarse)
		WW-17[d]	M	% coarse = 94.7
Everett Harbor	EVCHEM	SD-01	T,P,M,C	% coarse = 95.5
Eagle Harbor	EHCHEM	EH-02	M	% coarse = 89.3; DL
		EH-03	T,M	% coarse = 92.1; DL
Alki Pt.	ALKI	AP-04	M	% coarse = 95.9
Elliott Bay	TPPS3AB	EB-38 (3/15/82)	M	no obvious factors (13% coarse)
		EB-38 (7/15/82)	T,P,M	no obvious factors (38% coarse)
		WP-03 (7/15/82)	P	% coarse = 92.4; DL
		WP-08 (7/15/82)	M	% coarse = 92.6
		WP-16 (3/15/82)	P	no obvious factors (4% coarse)

[a] Taxa showing significant depressions. T = total taxa, P = Polychaeta, M = Mollusca, C = Crustacea.

[b] Percent coarse-grained material (i.e., ≥63 um).

[c] DL = Detection limits of at least one chemical exceeded 1988 benthic infauna AET for that chemical.

[d] Station predicted by 1986 benthic AET but not 1988 benthic AET.

[e] Station NS-08 is located near the Pier 91 naval dock. Tributyltin contamination is possible in this area but has not been tested.

high percentage (for example, ≥75 percent) of fine-grained sediment. By contrast, significant mortality was associated with less than 30 percent of the remaining stations. In the present study, 16 of the 29 impacted amphipod bioassay stations (55 percent) that were not predicted by AET were characterized by sediments having a relatively high percentage of fine-grained material. In the EIGHTBAY survey (Table 10), 81 percent of the impacted stations that were not predicted could be explained by this factor. A significant depression in benthic infaunal abundance was found at only one of these fine-grained stations (EBCHEM Station NS-08). However, adequate benthic infaunal data were not available at all amphipod bioassay stations (e.g., none of the stations in the EIGHTBAY survey).

A linear regression equation developed by DeWitt et al. (1988) has been suggested as a possible means of adjusting amphipod bioassay results to account for grain-size effects. DeWitt et al. (1988) developed a significant ($P \leq 0.001$) relationship that explained 29 percent of the variance in the relationship between *Rhepoxynius abronius* mortality and percent fine-grained material for sediments from reference areas of Puget Sound and Oregon. Stations with mean mortality greater than the 95 percent prediction limit for the regression relationship was suggested by DeWitt et al. (1988) as a guideline for defining toxic sediments. It is recommended that use of such a regression analysis be evaluated further as a means of identifying stations at which biological effects are potentially explained by alternative factors related to grain size distributions. A brief evaluation of whether application of the DeWitt regression equation accounts for the apparent amphipod mortality in 16 fine-grained stations listed in Table 10 was conducted. This evaluation suggested that grain size or related factors may have influenced toxicity at all but four of these stations (i.e., Stations EW-10, NS-08, SM-03, and BH-23).

Elevated detection limits precluded an assessment of whether one or more AET were exceeded at nearly all of the impacted amphipod bioassay stations that were not predicted. Aside from detection limit concerns or the possible presence of unmeasured chemicals, no obvious factors account for the apparent incorrect predictions at 12 of the 29 stations (41 percent). Most of these amphipod bioassay stations were found near Harbor Island, an area contaminated with a variety of chemical classes, suggesting that the biological effects could have resulted from unmeasured chemicals or from interactive effects among the numerous chemicals present at most of those stations.

Benthic Infauna Stations--The characteristics of benthic macroinvertebrate assemblages are influenced, in part, by sediment grain size distribution. Coarse-grained sediments may be suboptimal habitats for many benthic species because they are generally low in organic content and are indicative of high-energy environments. Thus, the food supply for benthic infauna may be limited, and organisms may have difficulty maintaining burrows, tubes, or position in the shifting sediment. In the present study, 16 of the 25 impacted benthic infauna stations (64 percent) that were not predicted by AET were characterized by sediments having a relatively high percentage (for example, ≥70 percent) of coarse-grained sediment. For example, the observed depression of benthic infaunal abundance at Station SD-01 in Everett Harbor is potentially attributable to the location of this station in a current-swept channel on the Snohomish River delta, rather than toxic effects. Elevated detection limits precluded an assessment of whether one or more AET were exceeded at nearly all of the impacted stations that were not predicted (Table 11).

Aside from potential concerns over detection limits, there was no other obvious factor that might explain apparent incorrect predictions at 9 of the 25 stations (36 percent). As discussed above for amphipod bioassay stations, most of these potentially

unexplained stations occurred in the Harbor Island area of the lower Duwamish River. Various dredging projects have occurred in this area and may have disturbed benthic assemblages sufficiently to result in depressions of major taxa. However, detection limits, unmeasured chemicals, or chemical synergism are the most likely factors at these Harbor Island stations, especially considering that four of the five incorrectly predicted stations at which both the amphipod bioassay and benthic infauna indicators were significant were from this area [i.e., Stations KG-11, NH-02, NH-11, and WW-08 from Beller et al. (1988)]. The fifth such station (Station NS-08) is located near Pier 91 in Elliott Bay. Potential (as yet untested) tributyltin contamination from historical naval ship operations may account for the substantial amphipod mortality and benthic infaunal depressions observed at this station.

Although unmeasured chemicals or other factors may account for some of the impacted stations that were not predicted by AET, sediment grain size appears to be a dominant factor at the majority of these stations. The fact that fine-grained sediments dominate such amphipod bioassay stations and that coarse-grained sediments dominate such benthic infauna stations reinforces this conclusion. Only 17 percent of the impacted amphipod bioassay stations that were not predicted had a relatively high percentage of coarse-grained material, and only 16 percent of the impacted benthic infauna stations that were not predicted had a relatively high percentage of fine-grained sediments. No correction factor is presently recommended for grain size, other than matching reference area sediments as closely as possible with sediments from the study site. The 1988 AET are efficient with respect to not predicting effects for stations at which grain size or associated factors may dominate over toxic effects.

3.3.2 Influence of Geographic Location on AET Reliability

Because AET are based on empirical data, the influence of sample location is important in evaluating the applicability of AET. The reliability of predictions for 1988 Puget Sound AET by survey and geographic area were summarized in Figure 5 and Table 8 (see Section 3.2). These AET are 100 percent efficient in predicting effects for the 1988 Puget Sound database (excluding chemical anomalies defined in Section 2.4.2; by including these anomalies, the 1988 amphipod AET are 88 percent efficient and the 1988 benthic infauna AET are 95 percent efficient). No obvious trends in sensitivity were observed for amphipod mortality and benthic infaunal abundance indicators in the major urban bay and Superfund surveys in Commencement Bay, Eagle Harbor, Elliott Bay, and Everett Harbor.

Nearly identical sensitivity was observed for both indicators in the Commencement Bay and Everett Harbor surveys conducted over two years apart in north and south Puget Sound. From 60 to 100 percent sensitivity was observed in all geographic areas for which all chemical classes represented by AET were analyzed in a survey, including tentatively identified organic compounds. In general, there were greater differences in sensitivity among surveys in the same location than among geographic areas sampled in Puget Sound. For example, 70 percent sensitivity was observed in amphipod bioassay samples from the Duwamish River in the recent Elliott Bay survey compared with 43-50 percent sensitivity in dredging surveys conducted as part of PSDDA in the same area (Table 8). The PSDDA surveys also analyzed fewer chemicals than were analyzed in the recent Elliott Bay survey. In addition, an unusually low sensitivity was observed for the amphipod bioassay AET when tested against the 48-station EIGHTBAY survey (Figure 5) although this survey included samples from the same urban areas in which high sensitivity was seen in other surveys. As discussed in the previous section, 81 percent of the impacted stations that were not predicted by AET in the EIGHTBAY survey are associated

with a high percentage of fine-grained material in the sediment, which may influence amphipod mortality. High detection limits for many organic compounds also exceeded AET for this survey (note that if chemicals were actually present at or near the high detection limits for these chemicals, the sensitivity of the AET predictions would approach 100 percent; however, detection limits have not been used to predict adverse effects for any survey).

As part of additional AET tests, an evaluation was made to determine whether AET based solely on data for one geographic area showed better predictive reliability when applied to the same area than when applied to different areas not included in the development of the AET. For this test, AET based on 16 stations from three surveys conducted in the Duwamish River (EBCHEM, DUWRIV1, and DUWRIV2) were calculated and used to predict biological effects for new PSDDA data collected in the Duwamish estuary, the recently completed PSEP study in Everett Harbor, and the Superfund studies in Commencement Bay.

As summarized in Table 12, there was 100 percent sensitivity and low efficiency (approximately 33 percent) in all areas evaluated. Hence, there was no discernable difference in the reliability of AET generated and applied in the same geographic area in comparison with the same AET applied in different geographic areas. The results also suggest that the sample size of 16 stations from the Duwamish River was too small or nonrepresentative of a range of contaminant conditions for multiple chemicals to generate efficient AET values (see following section).

3.3.3 Data Requirements to Derive Reliable AET

An evaluation was conducted to estimate the minimum number of stations needed to derive reliable AET. This effort was conducted by analyzing the reliability of AET developed from a series of randomly selected subsets of the updated SEDQUAL database for Puget Sound. Random subsets were selected in groups of 10, 30, and 50 stations for each biological indicator. The resulting AET calculated for each random subset were then used to predict effects for the appropriate biological indicator in the entire database (including the random subset so that the total number of stations used for predictions remained constant).

In addition to these random subsets of 10, 30, and 50 stations, the single station deletion test discussed previously (see Section 3.2) provides an estimate of the predictive reliability of the total database (e.g., 287 amphipod bioassay database). An alternative test was also conducted to approximate the predictive reliability of a data set equal to 95 percent of the size of the total database. This alternative test was conducted by randomly deleting 5 percent of the total number of stations, recalculating AET and then predicting effects on the deleted 5 percent of stations. These random deletions were repeated for a minimum of 8 trials before combining the results of all trials to determine sensitivity and efficiency. Three sets of 8 trials were conducted to achieve three estimates of sensitivity and efficiency for each biological indicator.

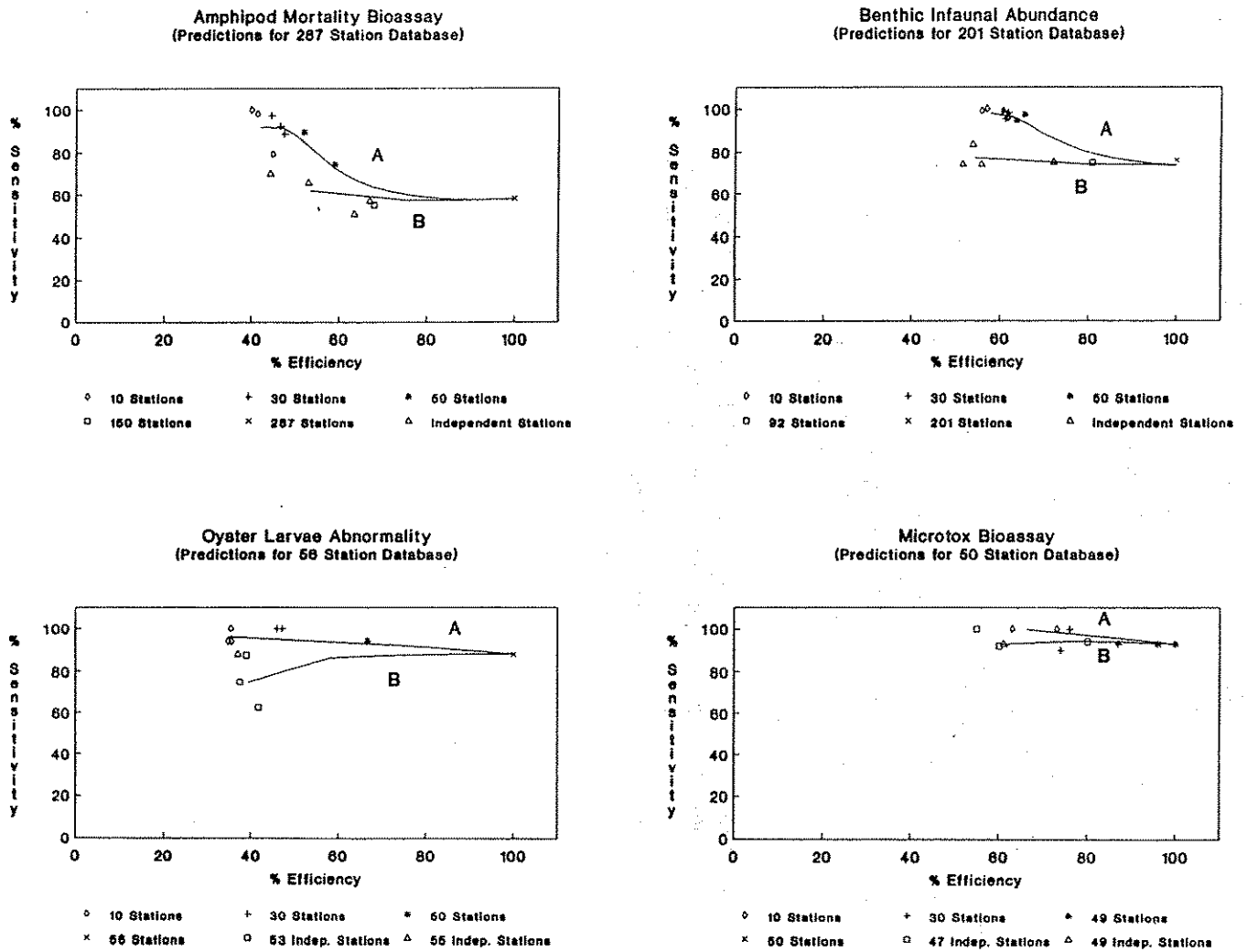
From these results, two different curves of sensitivity and efficiency as a function of sample size used to generate AET can be constructed. Such curves for each biological indicator are shown in Figure 8 and Table 13. Curve A for each biological indicator represents the change in sensitivity and efficiency in predicting effects based on random sets of 10, 30, 50 stations and the total number of stations in the SEDQUAL database. The sensitivity and efficiency of the 1986 amphipod AET (generated from 150 stations) and benthic infauna AET (generated from 94 stations) are also plotted.

TABLE 12. RELIABILITY OF INDEPENDENT "TEST" AET IN PREDICTING EFFECTS AT THREE GEOGRAPHIC AREAS[a]

Embayment	Survey	Sensitivity	Efficiency
Everett Harbor	EVCHEM	100% (7/7)	27% (7/26)
Commencement Bay	CBBLAIR and CBMSQS	100% (19/19)	35% (19/55)
Duwamish River	NOADW86[b]	100% (2/2)	33% (2/6)

[a] AET derived using EBCHEM data for Duwamish River stations (DR-) only. Duwamish River predictions were then made on a independent dataset (NOADW86). The Duwamish River is the only Puget Sound area for which multiple survey data were available for this comparison.

[b] Results exclude predictions made solely on sulfides. Sulfides data for NOADW86 survey are considered anomalous due to unusually high concentrations exceeding the AET by factors ranging from 3 to 163 at all impacted and nonimpacted stations. One nonimpacted sample contained sulfides at a high concentration of 18,000 mg/kg dry weight (95 times the sulfides AET).



Note: Independent station AET correspond to 5 percent, and single station deletion tests (see text and Table 13).

Figure 8. Reliability of AET for different sized data sets.

TABLE 13. RELIABILITY BY SIZE OF DATASET DETERMINING AET

Number of Stations	Sensitivity	Efficiency
Benthic AET vs. Benthic Infaunal Abundance Stations		
10 random stations	100% (108/108)	57% (108/189)
	99% (107/108)	55% (107/193)
	95% (103/108)	61% (103/167)
30 random stations	98% (106/108)	62% (106/171)
	95% (103/108)	61% (103/168)
	96% (104/108)	62% (104/167)
50 random stations	94% (102/108)	64% (102/160)
	97% (105/108)	65% (105/160)
	99% (107/108)	61% (107/176)
92 stations (1986 AET)	76% (82/108)	82% (82/100)
201 stations (1988 AET)	75% (81/108)	100% (81/81)[a]
191 independent stations (5% random deletion test)	74% (34/46)	56% (34/61)
	83% (35/42)	54% (35/65)
	74% (34/46)	52% (34/66)
200 independent stations (single station deletion test)	75% (81/108)	72% (81/112)
Amphipod AET vs. Amphipod Mortality Bioassay Stations		
10 random stations	100% (106/106)	40% (106/267)
	98% (104/106)	41% (104/253)
	79% (84/106)	44% (84/190)
30 random stations	89% (94/106)	47% (94/199)
	97% (103/106)	44% (103/232)
	92% (98/106)	46% (98/211)
50 random stations	90% (95/106)	52% (95/184)
	90% (95/106)	52% (95/184)
	75% (79/106)	59% (79/135)
150 stations (1986 AET)	55% (59/106)	69% (59/86)
287 stations (1988 AET)	58% (62/106)	100% (62/62)[a]
273 independent stations (5% random deletion test)	70% (23/33)	44% (23/52)
	66% (27/41)	53% (27/51)
	51% (26/51)	63% (26/41)
286 independent stations (single station deletion test)	57% (60/106)	67% (60/90)

TABLE 13. (Continued)

Number of Stations	Sensitivity	Efficiency
Microtox AET vs. Microtox Bioassay Stations		
10 random stations	100% (29/29)	73% (29/40)
	100% (29/29)	63% (29/46)
	100% (29/29)	63% (29/46)
30 random stations	90% (26/29)	74% (26/35)
	93% (27/29)	87% (27/31)
	100% (29/29)	76% (29/38)
49 random stations	93% (27/29)	100% (27/27)
	93% (27/29)	96% (27/28)
	93% (27/29)	100% (27/27)
50 stations	93% (27/29)	100% (27/27)[a]
47 independent stations (5% random deletion test)	94% (16/17)	80% (16/20)
	92% (12/13)	60% (12/20)
	100% (12/12)	55% (12/22)
49 independent stations (single station deletion test)	93% (27/29)	61% (27/44)
Oyster Larvae AET vs. Oyster Larvae Abnormality Stations		
10 random stations	100% (17/17)	35% (17/48)
	94% (16/17)	36% (16/45)
	94% (16/17)	35% (16/46)
30 random stations	100% (17/17)	47% (17/36)
	100% (17/17)	46% (17/37)
	100% (17/17)	46% (17/37)
50 random stations	94% (16/17)	67% (16/24)
	94% (16/17)	67% (16/24)
	94% (16/17)	67% (16/24)
56 stations	88% (15/17)	100% (15/15)[a]
53 independent stations (5% random deletion test)	63% (5/8)	42% (5/12)
	75% (6/8)	38% (6/16)
	88% (7/8)	39% (7/18)
55 independent stations (single station deletion test)	88% (15/17)	37% (15/41)

[a] By definition, efficiency is 100% because all Puget Sound stations were included in the calculation of these AET.

The stations used to generate AET for each data point on Curve A were also used in the prediction of biological effects on the total database.

Curve B for each biological indicator represents the change in sensitivity and efficiency in predicting effects using the 5 percent and single deletion tests, and the total number of stations in the SEDQUAL database. This latter curve represents predictions that are completely independent of the stations used to generate AET (except for the endpoint represented by the 1988 AET generated from the total database).

In general, sensitivity decreases as efficiency approaches 100 percent for randomly selected stations used to calculate AET on Curve A; the decline in sensitivity is greater when there is a large database (i.e., >100 samples for amphipod bioassay and benthic infauna) compared to a smaller database (i.e., approximately 50 samples for oyster larvae and Microtox bioassays). The trends for Curve A in Figure 8 suggest the following conclusions:

- The gain in efficiency as larger groups of stations are used to generate AET is always greater than the loss in sensitivity for every one of the biological indicators.

- Efficiency approaching 100 percent is not achieved until the entire Puget Sound database is used to calculate AET. Such an outcome is understandable given that only random selection of the exact group of stations uniquely setting AET values would yield 100 percent efficiency in predictions against the total database

- Sensitivity generally is high (although efficiency may be low) when small numbers of stations are used to generate AET for multiple contaminants. This outcome suggests that small numbers of stations (e.g., $\leq$30 stations) typically generate environmentally protective AET for problem identification, but that confirmation of the potential problems identified is clearly warranted.

Based upon the results for Curve A compiled in Table 13, two additional conclusions can be made concerning efficiency:

- The number of correctly predicted impacts remains relatively constant or decreases only slightly as larger numbers of randomly selected stations are used to generate AET.

- The number of incorrectly predicted impacts tends to decrease (hence, efficiency tends to improve) when a larger number of stations are used to generate AET. This result could imply that for Puget Sound the increase in the size of the AET database has tended to improve estimates of sediment quality values by increasing AET for chemicals that were originally established too low because of a nonrepresentative range of concentration data.

Examination of results for Curve B (Figure 8) indicate that sensitivity is relatively unchanged but efficiency increases as larger numbers of independent stations are used to calculate AET for predictions of adverse biological effects. It should be noted that efficiency is expected to be less than 100 percent in these tests because the random deletion of any of the stations uniquely setting an AET value in the SEDQUAL database

will result in a lowering of efficiency when the resulting AET are used to predict effects at that station.

The results also suggest that a biased sampling plan should always be used when initially building the database (i.e., always including samples from known contaminated areas) to ensure that a wide range of contaminant concentrations is represented rather than a completely random sampling of the environment. Accordingly, biased sampling should be increasingly important to ensure reasonable efficiency of AET when generated from small numbers of samples. It is also important that a number of medium to high concentration stations be included, such that several data points will lie within the range of the AET that is defined.

The variation of AET for individual chemicals for the random trials of 10, 30, and 50 stations was also examined for potential trends. In general, AET based on random groups of 10 to 30 stations were highly variable or missing (i.e., could not be calculated either because all data for the chemical were undetected or were "tentative" because there was no impacted station with a higher concentration than the AET). Selection of random groups of 50 stations from the entire Puget Sound database generally resulted in less variable AET than for 10- or 30-station groups, and 50 stations randomly sampled from only the contaminated urban embayments resulted in even more stable AET. These results, the sensitivity/efficiency results summarized in Figure 8, and the predictions for different geographic areas based on test AET calculated for 16 stations from a single area in the Duwamish River (see previous Section 3.3.2) suggest the following strategy for developing reliable AET:

- Sample more than 30, and preferably at least 50 stations.
- Bias the positioning of stations to ensure sampling of a variety of contaminant sources (e.g., an urban environment impacted by multiple contaminant sources, and preferably representative of the range of geographic areas to which the AET will be applied) over a range of contaminant concentrations (e.g., a factor of 10-100 is suggested but there are no definitive data to recommend a critical factor)
- Conduct chemical tests for a wide range of chemical classes (see Section 3.3.4) using low-level chemical analyses (low detection limits are especially important for trace organic analyses; <100 ppb detection limits appear to be necessary, especially for contaminants other than PAH).

If small numbers (i.e, <50) of randomly selected stations are used to develop AET, the resulting values are expected to be sensitive but not highly efficient predictors of impacted stations in independent tests at contaminated sites.

3.3.4 Identification of Reliable Subsets of Chemical AET

Empirical sediment quality values, such as AET, are based on relating chemical concentrations with the occurrence of biological effects. An AET derived for a chemical suggests but does not prove a toxic cause-effect relationship for that chemical. Because chemicals occur in mixtures in the environment, and other factors may also influence biological indicators, it is possible that some AET do not contribute to the overall sensitivity of the method in identifying problem sediments or are otherwise redundant with other AET. The purpose of the evaluations in this section is to determine whether using a subset of chemical AET can achieve the same sensitivity as the entire set of

AET, not to draw conclusions concerning possible cause-effect relationships. A sensitive and efficient subset of AET could provide a cost-effective tool for managing contaminated sediments by reducing the number of chemicals that require testing or by reducing the amount of data that must be reviewed by decision-makers. Reliance on this evaluation for sediment management must be carefully considered, as the results for Puget Sound data now in SEDQUAL may not necessarily be applicable to other data sets.

This evaluation was conducted using 1988 AET for the amphipod bioassay and benthic infauna indicators based on available biological effects data in the SEDQUAL database. It is important to note that not all chemicals were measured at all stations, therefore, interpretations on the importance of particular chemical groups could be biased against those chemicals that have not been routinely measured. The AET used for this analysis were based on the following data and considerations:

- Data for a total of 214 chemicals, chemical groups (e.g., HPAH), or conventional variables (e.g., sulfides) were reported for one or more sampling stations in SEDQUAL. Ninety two (92) of these chemicals were infrequently detected or reported in Puget Sound (i.e., $\leq$5 detected values; Table 14) and were deleted from calculation of AET. The remaining 122 chemicals that were detected in >5 samples were measured at >10 (and in some cases several hundred) stations in the database.

- Definitive AET (see Section 2.2) could be developed for 104 of these 122 chemicals (87 amphipod bioassay AET and 96 benthic AET). The remaining chemicals are potentially not associated with adverse biological effects over the Puget Sound concentration range represented in SEDQUAL.

- Of the 104 AET, 31 were not initially considered because the AET (a) were exceeded at $\leq$5 impacted stations and other chemicals still predicted those stations; (b) were based on chemicals that were tentatively identified organic compounds for which specific structures could not be assigned (e.g., an unidentified alkanol); (c) were based on chemicals exhibiting a narrow concentration range in the SEDQUAL database for Puget Sound (i.e., selenium, barium, trichloroethene); (d) were based on chemicals that were nearly always detected with more frequently detected contaminants (e.g., the AET for PCBs nearly always was exceeded for stations at which the AET for the less frequently detected DDT isomers were exceeded; the AET for dehydroabietic acid was always exceeded at stations for which AET for other resin acids, chlorinated guaiacols, and chlorinated phenols associated with pulp mill wastes were exceeded; the AET for the EPA priority pollutant hexachlorobutadiene was nearly always exceeded at stations for which AET for tentatively identified tri-, tetra-, and pentachlorobutadiene isomers were exceeded). As is discussed later in this section, 7 of these 31 AET contribute slightly to the overall sensitivity of the 1988 AET.

Repeated reliability evaluations were conducted on the 73 remaining AET, which were categorized into the following 13 chemical groups using best professional judgement:

- 5 conventional variables (total organic carbon, sulfides, total volatile solids, oil & grease, percent fine-grained material)

- 9 metals (arsenic, cadmium, chromium, copper, lead, mercury, nickel, silver, zinc, and excluding antimony, beryllium, selenium, and thallium)

TABLE 14. INFREQUENTLY DETECTED CHEMICALS NOT CONTRIBUTING TO SENSITIVITY OF 1988 AET

Chemical	Detection Frequency (Amphipod Surveys)	Detection Frequency (Benthic Surveys)
Volatiles		
1,1-dichloroethane	1/27	1/27
1,1-dichloroethene	1/37	1/37
1,2-dichloroethane	0/37	0/37
1,2-dichloroethene	2/37	2/37
1,2-dichloropropane	0/37	0/37
1,1,1-trichloroethane	0/10	0/10
1,1,2-trichloroethane	0/37	0/37
1,1,2,2-tetrachloroethane	0/37	0/37
2-butanone	0/10	0/10
4-methyl-2-pentanone	0/10	0/10
2-hexanone	0/10	0/10
chloroethylvinyl ether	0/27	0/27
2-chloroethylvinyl ether	0/10	0/10
acetone	1/10	1/10
acrolein	0/27	0/27
acrylonitrile	0/27	0/27
benzene	0/37	0/37
bromodichloromethane	0/37	0/37
bromomethane	0/37	0/37
bromoform	0/37	0/37
carbon tetrachloride	0/37	0/37
cis 1,3-dichloropropene	0/27	0/27
cis 1,3-dichloropropane	0/10	0/10
chloroform	2/106	2/58
chlorobenzene	2/37	2/37
chloroethane	0/37	0/37
chloromethane	0/37	0/37
carbon disulfide	0/10	0/10
dibromochloromethane	0/37	0/37
methylene chloride	0/10	0/10
nitrobenzene	0/41	0/41
styrene	0/10	0/10
trans-1,3-dichloropropene	0/37	0/37
toluene	3/37	3/37
vinyl acetate	0/10	0/10
vinyl chloride	0/37	0/37

TABLE 14. (Continued)

Chemical	Detection Frequency (Amphipod Surveys)	Detection Frequency (Benthic Surveys)
Oxygenated Semivolatiles		
bis(2-chloroethyl)ether	1/41	1/41
bis(2-chloroisopropyl)ether	0/31	0/31
bis(2-chloroethoxy) methane	0/41	0/41
4-bromophenylphenylether	0/31	
4-chlorophenylphenylether	0/41	0/41
Pesticides		
aldrin	3/190	3/190
chlordane	1/189	1/189
dieldrin	1/189	1/189
endosulfan-I	0/31	0/31
endosulfan-II	0/31	0/31
endosylfan sulfate	0/31	0/31
endrin	1/134	1/134
endrin ketone	0/31	0/31
endrin aldehyde	1/103	1/103
heptachlor epoxide	0/31	0/31
heptachlor	2/190	2/190
alpha-hexachlorocyclohexane	0/134	0/134
beta-hexachlorocyclohexane	0/134	0/134
delta-hexachlorocyclohexane	1/134	1/134
gamma-hexachlorocyclohexane	1/190	1/190
isophorone	4/249	4/249
methoxychlor	0/31	0/31
toxaphene	0/31	0/31
Resin Acids		
14-chlorodehydroabietic acid	3/11	3/11
dichlorodehydroabietic acid	2/11	2/11
isopimaric acid	5/11	5/11
neoabietic acid	2/11	2/11
sandaracopimaric acid	5/11	5/11
Phenols		
2-chlorophenol	5/134	5/134
2-nitrophenol	0/41	0/41
4-nitrophenol	0/41	0/41
4,6-dinitrocresol	0/41	0/41
2,4-dinitrophenol	0/41	0/41
4-chloro-3-methylphenol	0/144	0/144
3,4,5-trichloroguaiacol	1/11	1/11
4,5,6-trichloroguaiacol	1/11	1/11
tetrachloroguaiacol	2/11	2/11
2,3,4,6-tetrachlorophenol	2/11	2/11

TABLE 14. (Continued)

Chemical	Detection Frequency (Amphipod Surveys)	Detection Frequency (Benthic Surveys)
Organic Bases		
aniline	1/52	1/52
4-chloroaniline	0/31	0/31
2-nitroaniline	0/31	0/31
3-nitroaniline	0/31	0/31
4-nitroaniline	0/31	0/31
2,4-dinitrotoluene	0/41	0/41
2,6-dinitrotoluene	0/146	0/128
1,2-diphenylhydrazine	3/97	4/88
N-nitroso-di-n-propyl amine	0/31	0/31
3,3'-dichlorobenzidine	0/31	0/31
Miscellaneous Neutrals		
1-methylnaphthalene	5/5	5/5
2,5-dimethylnaphthalene	5/5	5/5
2-chloronaphthalene	0/144	0/144
hexachlorocyclopentadiene	0/46	0/31
hexachloroethane	2/137	2/128
perylene	5/5	5/5
Other Chemicals		
tributyltin[a]	1/2	1/2

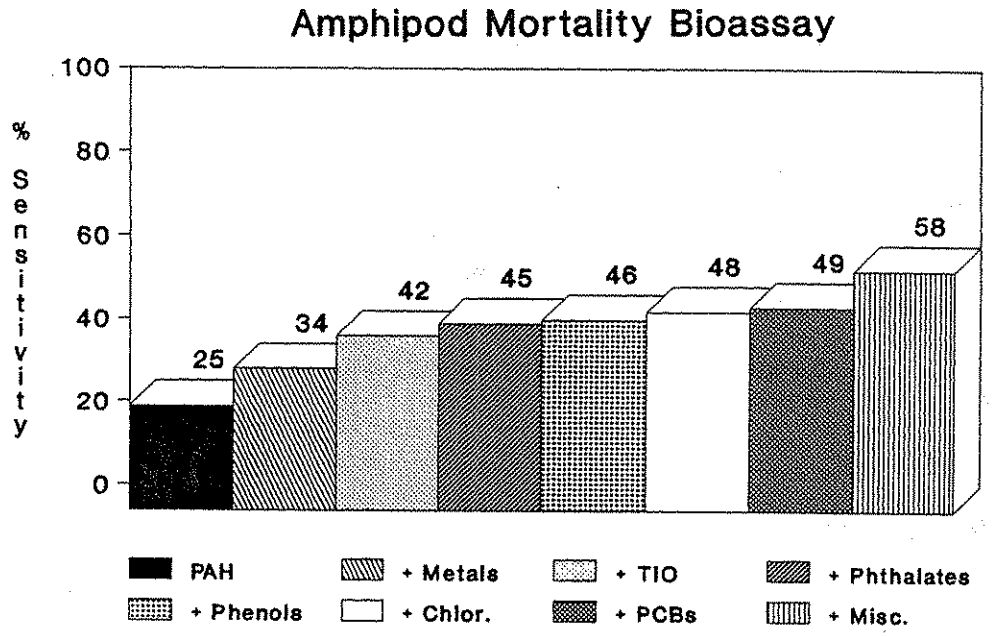
[a] Tributyltin analyses have only recently been conducted at selected sites. Tributyltin is known to be toxic at low concentrations and likely could contribute to observed effects at other sites at which tributyltin analyses have not yet been conducted. Therefore, unlike most of the other chemicals in this table, acquisition of additional AET data for tributyltin is recommended.

- 16 individual PAH and group sums of LPAH and HPAH
- Total PCBs
- 5 phthalate esters [bis(2-ethylhexyl)phthalate, butyl benzyl phthalate, diethyl phthalate, dimethyl phthalate, di-n-octyl phthalate]
- Phenol and 3 alkylated phenol compounds (2-methylphenol, 4-methylphenol, 2,4-dimethylphenol)
- Pentachlorophenol (in general, the distribution of this chemical does not covary with those of other chlorinated phenolic compounds)
- Dehydroabietic acid [this resin acid was an effective representative of resin acids and chlorinated phenolic compounds (excluding pentachlorophenol), most of which were measured only recently in the Everett Harbor PSEP study]
- 2 miscellaneous polar compounds (benzoic acid and benzyl alcohol)
- 5 miscellaneous chlorinated neutral compounds (1,2-dichlorobenzene, 1,4-dichlorobenzene, 1,2,4-trichlorobenzene, hexachlorobenzene, hexachlorobutadiene)
- 2 miscellaneous base/neutral compounds (N-nitrosodiphenylamine, dibenzofuran)
- 17 tentatively identified organic compounds [this group was not subdivided according to chemical class and contained a variety of compound types including polar compounds (fatty acids, fatty acid methyl esters, 2-methoxyphenol, campesterol, coprostanol) and neutral compounds (miscellaneous alkylated PAH, diterpenoid hydrocarbons, carbazole, 1,2,4-trithiolane, dibenzothiophene)
- 3 volatile organic compounds (tetrachloroethene, ethylbenzene, and total xylenes, which were the only volatile compounds detected in >5 sediment samples).

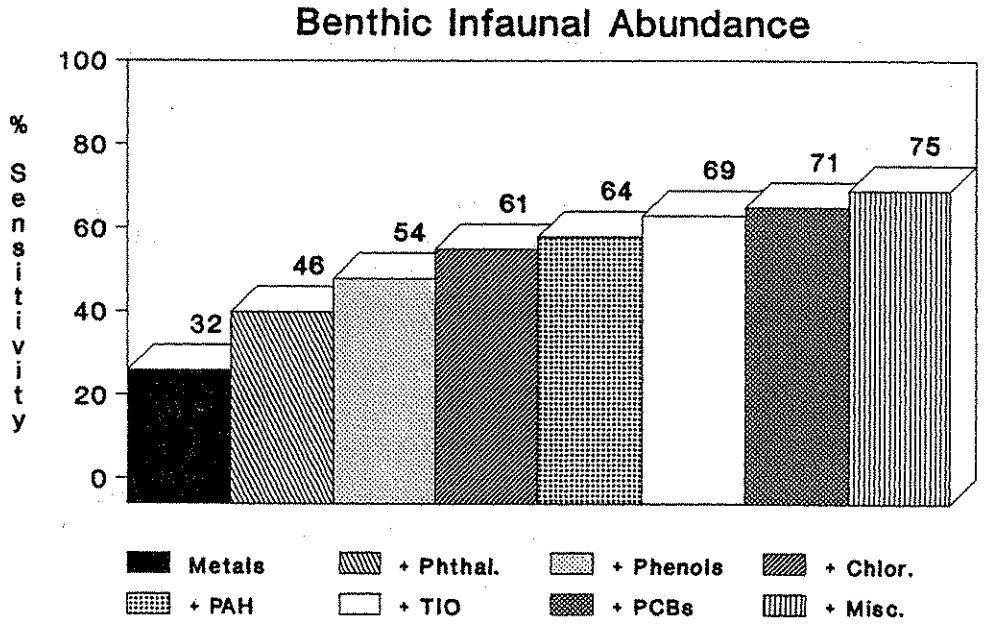
The AET group uniquely accounting for the largest percentage of the overall sensitivity of Puget Sound AET was determined for each biological indicator. This evaluation was conducted by combining all AET groups and then sequentially deleting individual groups of AET to determine the numbers of stations that were uniquely predicted by each AET group (i.e., such stations are no longer predicted as impacted when the particular AET group is deleted from the overall Puget Sound AET).

The overall sensitivity of 1988 amphipod bioassay AET using all chemicals is 58 percent (see Table 9). PAH uniquely account for approximately 11 percent sensitivity for amphipod bioassay stations, the largest percentage for any of the groups of amphipod bioassay AET. Metals uniquely account for approximately 8 percent sensitivity for benthic infauna stations, the largest percentage for any of the groups of benthic infauna AET.

In addition, results plotted in Figure 9 identify the incremental increase in sensitivity determined by sequentially combining the above groups of AET, starting with the group



Sensitivity of 1988 Amphipod AET = 58%
(Efficiency = 100% by constraint;
see NOTE in Table 7)



Sensitivity of 1988 Benthic AET = 75%
(Efficiency = 100% by constraint;
see NOTE in Table 7)

Figure 9. Incremental increases in sensitivity of 1988 amphipod and benthic infauna AET due to sequential addition of data for various chemical groups.

that uniquely accounts for the largest percentage of the overall sensitivity. As shown in Figure 9, stations correctly predicted using PAH AET account for 25 percent sensitivity for amphipod bioassay stations in all Puget Sound surveys. (This percentage includes all stations that are uniquely predicted by PAH AET and those at which PAH and other AET groups redundantly predict adverse effects.) The combination of AET for PAH and metals accounted for an additional 9 percent (total of 34 percent) sensitivity. The combination of PAH, metals, and tentatively identified organic compounds accounted for an additional 8 percent (total of 42 percent) sensitivity. All of the sensitivity (58 percent) was accounted for by seven groups of AET plus the seven miscellaneous chemicals listed in Table 15. The remaining six groups of AET listed above do not uniquely contribute to the overall sensitivity of amphipod bioassay AET.

Results for benthic infauna AET are also shown in Figure 9. The overall sensitivity of 1988 benthic infauna AET using all chemicals is 75 percent (see Table 9). Stations correctly predicted using metals AET account for 32 percent sensitivity for benthic infauna stations. (This percentage includes all stations that are uniquely predicted by metals AET and those at which metals and other AET groups redundantly predict adverse benthic effects.) The combination of metals and phthalate esters accounted for an additional 14 percent (total of 46 percent) sensitivity. All of the sensitivity (75 percent) was accounted for by the same seven groups of AET as determined for amphipod bioassay AET, plus antimony (see Table 15). Several conclusions can be drawn from these analyses, some of which have implications for prioritizing laboratory analyses:

- The apparent relative importance of different chemical groups differs for the two different biological indicators; metals uniquely account for the largest portion of predicted *in situ* benthic infaunal impacts, while PAH uniquely account for the largest portion of predicted impacts according to the acute lethal amphipod bioassay

- Predictions based on AET for multiple chemical classes (including inorganic and organic contaminants) are required to account for a high proportion of effects measured by either the amphipod bioassay or benthic infaunal indicators in Puget Sound (i.e., no one chemical or chemical class accounts for more than approximately half of the total sensitivity attained using all chemicals)

- The following chemical groups do not uniquely account for predicted impacts for either the amphipod bioassay or benthic infauna AET:

  - Standard conventional variables (e.g., TOC, TVS, sulfides, grain size, oil & grease) (measurements of TOC, sulfides, and grain size are required for normalizing chemical concentrations or interpreting biological conditions but TVS and oil & grease measurements are not)

  - Volatile organic compounds, except trichloroethene (for two impacted amphipod bioassay stations from the Duwamish River; see Table 15)

  - Miscellaneous organic compounds including pentachlorophenol, resin acids, organic bases (e.g., N-nitrosodiphenylamine), and several Hazardous Substance List compounds including 2-methylnaphthalene, dibenzofuran, benzoic acid, and benzyl alcohol (low detection limits for the latter two compounds are sometimes difficult to attain but may not be essential)

TABLE 15. MISCELLANEOUS VARIABLES CONTRIBUTING TO THE SENSITIVITY OF 1988 AET

Chemical	Overall Detection Frequency	Biological Indicator	Survey	Station
Pentachlorophenol	24/231	AMPT[a]	EBCHEM	EW-16
2,3,5-Trimethylnaphthalene	5/5	AMPT	CBBLAIR	B15
Total chlorinated benzenes	7/103	AMPT	EBCHEM	DR-15
Total nitrogen	190/190	AMPT	CBMSQS	SP-16
Trichloroethene[b]	6/106	AMPT	EBCHEM EBCHEM	DR-13 KG-05
Thallium	46/114	AMPT	EIGHTBAY EIGHTBAY	EV-01 EV-05
p,p'-DDE[c]	25/247	AMPT	EBCHEM	DR-10
- - -	- - -	- - -	- - -	- - -
Antimony	206/221	BENA[d]	EBCHEM EBCHEM EBCHEM	WW-16 WW-18 WW-20

[a] The prediction of impacts at the nine amphipod stations shown account for a miscellaneous 9 percent of the total 58 percent sensitivity for the revised amphipod AET; the remaining 49 percent sensitivity is accounted for by classes of chemicals shown in Figure 9

[b] The amphipod AET for trans-1,2-dichloroethene is also exceeded at Station KG-05.

[c] The amphipod AET for oil & grease is also exceeded at Station DR-10.

[d] Prediction of impacts at the three benthic stations shown account for a miscellaneous 3 percent of the total 75 percent sensitivity for the updated benthic AET; the remaining 69 percent sensitivity is accounted for by classes of chemicals shown in Figure 9. Antimony was not included with other metals (a major predictive class of chemicals for benthic AET) because of analytical concerns over the reliability of the antimony data.

- PCB analyses uniquely account for 1-2 percent of the overall sensitivity of amphipod bioassay or benthic infauna AET; additional analysis of EPA priority pollutant pesticides accounts for less than 1 percent of the sensitivity of amphipod bioassay AET and for none of the sensitivity of benthic infauna AET. PCB analyses are still recommended, in part because of their importance in estimating potential human health effects and because PCBs are more important contributors to the sensitivity of other biological indicators (e.g., Microtox bioassay).

Based on these results, chemical analyses for volatile organic compounds do not appear to be routinely warranted. However, analyses limited to only EPA priority pollutants will not account for all of the predicted impacts, including areas contaminated with alkyl-substituted phenols (especially 4-methylphenol, which is a Hazardous Substance List compound but not an EPA priority pollutant), resin acids, or chlorinated phenols associated with pulp mill wastes. Analyses for metals, nonpolar organic compounds (especially PAH, chlorinated benzenes, hexachlorobutadiene, and phthalates), and organic acids (especially alkylated phenols) are required to account for a high percentage of the sensitivity of Puget Sound AET. In addition, tentatively identified organic compounds in the current Puget Sound database uniquely account for approximately 5-6 percent of the overall sensitivity of Puget Sound AET.

3.4 APPLICATION OF AET TO IDENTIFY POTENTIAL PROBLEM AREAS

Substantially more chemical data exist for different areas in Puget Sound than results for biological indicators. AET were applied to available chemical data in the Puget Sound database to identify potential problem areas, including areas in which only partial biological data are available to confirm predictions of biological effects.

Predictions of impacted stations were made for the approximately 300 stations in 11 surveys (Table 5) and for all four biological indicators represented by AET (benthic infauna abundance, and amphipod mortality, oyster larvae abnormality, and Microtox bioassays). These are the only predictions in this report that are made for stations at which confirming biological data were not always available. Therefore the predictions are, in some cases, tentative and can be confirmed by biological testing. Confirming data for all four of biological indicators were available only at stations in the CBMSQS survey in Commencement Bay. Maps identifying the location of predicted adverse effects are shown in Figures D-1 through D-13 in Appendix D.

During review of the draft report by the Puget Sound Sediment Criteria Workgroup, it was suggested that possible additive effects of chemicals at a station be accounted for by the cumulative ratio of the concentration of each sediment contaminant to its AET value (R. Swartz, personal communication). Stations exhibiting cumulative AET ratios of >1 would be predicted to exhibit adverse effects under an additive model of chemical toxicity. Adding ratios of individual contaminant concentrations to AET values implicitly assumes that each ratio is independent. However, individual AET determined from environmental samples that contain complex mixtures of chemicals presumably reflect some influence of chemical interactions (see Appendix A). To acknowledge potential "double counting" of interactive effects, it is recommended that such cumulative AET ratios be evaluated primarily for ranking problem stations at which at least one chemical concentration exceeds an AET for a particular biological indicator (i.e., ranking problem stations already predicted by individual AET values). For ranking, both the magnitude of AET exceedance and the number of AET exceedances are important in establishing a preponderance of evidence for predicting adverse biological effects.

4.0 SUMMARY

Evaluations conducted for this PSEP study indicate that 1988 AET values are generally reliable predictors of adverse biological effects after recalculating the 1986 AET values by incorporating new data from Eagle Harbor, Elliott Bay, and Everett Harbor. Of the 201 benthic infauna stations and 287 amphipod bioassay stations evaluated for 13 Puget Sound embayments, approximately 85 percent (174 stations and 243 stations, respectively) are in accordance with the predictions of the 1988 AET values for these indicators (i.e., they do not exhibit adverse effects at chemical concentrations less than the AET values, and do exhibit adverse effects at chemical concentrations above the AET values). One or the other of these biological indicators is still in accordance with predictions of the respective AET values at 23 of the 28 (82 percent) stations at which an impact was not predicted and data for both benthic infauna and amphipod bioassays were available. Because the 1988 AET are based on data for all available nonimpacted stations, none of the 91 nonimpacted benthic stations or 177 nonimpacted amphipod bioassay stations were mispredicted.

1986 AET values (dry weight normalization) are less efficient but are of similar sensitivity to the 1988 AET (dry weight normalization) and so have somewhat lower overall reliability. Of the 201 benthic infauna stations and 287 amphipod bioassay stations evaluated, approximately 75 percent (157 stations and 213 stations, respectively) are in accordance with the predictions of the 1986 AET values for these indicators. AET for the oyster larvae and Microtox bioassays have not been updated because new data using comparable bioassay protocols were not available (data are only available for Commencement Bay for these two indicators). Approximately 96 percent of both the existing 50-station oyster larvae database and the 50-station Microtox database are in accordance with the predictions of the 1986 values for these indicators.

TOC-normalized AET did not appear to offer an advantage in predictive success over dry weight-normalized AET. Specifically, for 1988 AET calculated for amphipod and benthic infauna indicators, use of TOC-normalized AET for nonionic organic compounds and dry weight-normalized AET for the remaining chemicals (metals and ionizable organic compounds) yielded approximately equivalent reliability results when compared with predictions by dry weight-normalized AET for all chemicals.

Physical characteristics (e.g., grain size, habitat exposure) and high detection limits (for which AET predictions are not made), or unanalyzed chemicals are the most likely factors accounting for the 29 amphipod and 24 benthic infauna stations that were not in accordance with predictions based on AET values (Section 3.3.1). High detection limits for at least one chemical at almost all of these stations prevented a complete comparison to AET values. A relatively high percentage ($\geq$75 percent) of fine-grained sediment was found at 55 percent of the incorrectly predicted amphipod bioassay stations (81 percent of the incorrectly predicted stations in the EIGHTBAY survey); such conditions are less favorable for the test organisms. A significant depression in benthic infaunal abundance was found at one of the two fine-grained sediments at which data for both biological indicators were available. These potential physical factors increase the uncertainty with which adverse biological effects can be identified. However, for the existing database, the 1988 AET are efficient with respect to not predicting effects for stations at which grain size or related factors may dominate over toxic effects.

A relatively high percentage (≥70 percent) of coarse-grained sediment was found at 64 percent of the incorrectly predicted benthic infauna stations and in some cases these samples were collected in areas of relatively high-wave intensity compared to reference conditions. Both conditions are expected to result in reduced abundances of organisms that may not have been adequately controlled for by comparisons to reference area samples. Significant mortality was observed in the amphipod bioassay at two of the eight coarse-grained sediment stations at which data for both biological indicators were available.

Detection limits, unmeasured chemicals, or unaccounted chemical interactions are important considerations at the five stations at which both significant amphipod mortality and significant depressions in abundance of benthic infauna occurred but no AET were exceeded. These stations are of concern because AET for both indicators failed to predict the observed effects and grain-size conditions were generally not considered to be a factor (these five stations constitute 3 percent of the 154 stations at which both amphipod bioassay and benthic infaunal data are available for comparisons). The five stations are from the PSEP Elliott Bay survey, and four of the five are located near Harbor Island in the lower Duwamish River, a federal Superfund site contaminated by a variety of organic and inorganic chemicals. The fifth station is located near Pier 91 in Elliott Bay; untested but potential tributyltin contamination from historical naval ship operations should be considered as a possible explanation for the substantial adverse effects observed at this station.

In general, there was little evidence to suggest that the reliability of AET was affected by geographic location. The 1988 AET are 100 percent efficient in predicting effects in areas of Puget Sound represented in the database (excluding chemically anomalous stations at which relatively high chemical concentrations but no adverse effects were observed; see Section 2.4.3). After including chemical anomalies in predictions, the 1988 amphipod AET are 88 percent efficient and the 1988 benthic infauna AET are 96 percent efficient for a total database of 295 and 205 stations, respectively. High and nearly identical sensitivity was observed for both amphipod bioassay and benthic infaunal abundance indicators in the Commencement Bay and Everett Harbor surveys conducted over two years apart in north and south Puget Sound. In general there were greater differences in sensitivity among individual surveys in the same location than among geographic areas. The lowest sensitivity (e.g., 17 percent for amphipod bioassay results in the EIGHTBAY study) was associated with surveys for which potential grain-size effects, high detection limits, or lack of data for all chemical classes were major factors.

From 60-100 percent sensitivity was observed in all geographic areas for which all chemical classes represented by AET were analyzed in a survey (including tentatively identified organic compounds). In a separate test, AET were generated solely from upper Duwamish River stations sampled in two surveys and compared with a later independent survey in the same area as well as with data from Everett Harbor and Commencement Bay. High (100 percent) sensitivity and low (31-35 percent) efficiency was attained in each of these comparisons. Hence, there was no evidence of differences in reliability among geographic regions although the test results indicated that a sample size of 16 stations from one geographic area was too small or nonrepresentative of a range of contaminant conditions to generate efficient AET values.

Based on these results and further tests using random selection of stations from the entire database (Section 3.3.3), it is recommended that the following strategy be used to develop reliable AET: 1) Collect chemical and biological effects data from preferably more than 50 stations; 2) Bias the positioning of stations to ensure sampling

of a variety of contaminant sources (e.g., an urban environment impacted by multiple contaminant sources, and preferably representative of the range of geographic areas to which the AET will be applied) over a range of contaminant concentrations (preferably over at least one to two orders of magnitude); 3) Conduct chemical tests for a wide range of chemical classes and ensure that <100 ppb detection limits (lower if possible) are attained for organic compounds (metals detection limits do not appear to be a problem). In general, small numbers of randomly selected stations yield AET that are sensitive (>80-100 percent) but not highly efficient (40-60 percent).

The sensitivity of 1988 AET is attributable to several chemical classes. No one chemical or chemical class (e.g., PAH, metals, phenolic compounds) accounts for more than half of the sensitivity that is attained using AET for all chemicals. However, a number of chemicals do not contribute to the sensitivity of the 1988 AET because they are infrequently detected or reported in Puget Sound (Table 14). In addition, several chemical variables do not appear to uniquely account for predicted impacts for either the amphipod bioassay or benthic infauna indicators, including standard conventional variables (TOC, TVS, sulfides, grain size, oil & grease), volatile organic compounds (except potentially trichloroethene in the Duwamish River), and several miscellaneous compounds (e.g., pentachlorophenol, resin acids, organic bases, dibenzofuran, benzoic acid, benzyl alcohol). Routine analysis for volatile organic compounds may not be warranted based on these results, which would represent a cost savings because these compounds require separate analysis.

The chemical groups that contribute most substantially to the sensitivity of the 1988 AET for amphipod bioassay and benthic infauna indicators include PAH, metals (excluding beryllium, selenium, barium, and naturally occurring iron or manganese), phenols, phthalate esters (not necessarily the few phthalate esters that are common laboratory contaminants), chlorinated neutral compounds (e.g., chlorinated benzenes, butadienes, and PCBs), and tentatively identified compounds (including PAH-related compounds and miscellaneous oxygenated compounds). An evaluation of the incremental increase in sensitivity by adding these chemical groups (Figure 9) indicated that PAH and metals were the two most important for identifying impacts in the Puget Sound database. However, these groups are also two of the most commonly measured variables, which may bias the results against other compound groups that have not been as widely measured at low detection limits (e.g., alkyl- and chlorinated phenols or chlorinated benzenes).

Finally, based on AET for the four available biological indicators (amphipod, oyster larvae, and Microtox bioassays, and benthic infaunal abundance), maps were prepared showing the location of predicted impacts in Puget Sound using most of the 334-station database. For ranking these stations, both the magnitude of AET exceedance and the number of AET exceedances are important in establishing a preponderance of evidence for predicting adverse biological effects. Comparison of the cumulative ratio of the concentration of each sediment contaminant to its AET value was suggested as a possible way to rank these stations that conforms to an additive model of chemical toxicity.

REFERENCES

Adams, W.J., R.A. Kimerle, and R.G. Mosher. 1985. Aquatic safety assessment of chemicals sorbed to sediments. pp. 429-453. In: Aquatic Toxicology and Hazard Assessment: Seventh Symposium. American Society for Testing and Materials, Philadelphia, PA.

Barrick, R., S. Becker, D. Weston, and T. Ginn. 1985. Commencement Bay nearshore/ tideflats remedial investigation. Final Report. Prepared by Tetra Tech, Inc. for Washington Department of Ecology and U.S. Environmental Protection Agency. Tetra Tech, Inc., Bellevue, WA.

Barrick, R., H. Beller, and M. Meredith. 1986. Eagle Harbor Preliminary Investigation. Final Report. Prepared by Tetra Tech, Inc. for Black & Veatch/Washington Department of Ecology. Tetra Tech, Inc., Bellevue, WA.

Barrick, R., L. Brown, and S. Becker. 1988. Sediment quality values refinement: Volume II; Evaluation of PSDDA sediment quality values. Prepared by PTI Environmental Services for Tetra Tech/U.S. Environmental Protection Agency Region 10, Office of Puget Sound. PTI Environmental Services, Bellevue, WA.

Battelle 1985. Detailed chemical and biological analyses of selected sediment from Puget Sound. Draft Final Report. U.S. Environmental Protection Agency Region 10, Seattle, WA. 300 pp.

Beller, H., R. Barrick, and S. Becker. 1986. Development of sediment quality values for Puget Sound. Prepared by Tetra Tech, Inc. for Resource Planning Associates/U.S. Army Corps of Engineers, Seattle District for the Puget Sound Dredged Disposal Analysis program. Tetra Tech, Inc., Bellevue, WA.

Beller, H., R. Pastorok, S. Becker, G. Braun, G. Bilyard, and P. Chapman. 1988a. Elliott Bay action program: analysis of toxic problem areas. Final Report. Prepared for U.S. Environmental Protection Agency Region 10, Office of Puget Sound. PTI Environmental Services and Tetra Tech, Inc., Bellevue, WA. July, 1988.

Carter C.W., and I.H. Suffet. 1985. Quantitative measurements of pollutant binding to dissolved humic materials compared with bulk properties of humic materials. Org. Geochem. 8:145-146.

Chan, S., M. Schiewe, and D. Brown. 1985. Analysis of sediment samples for U.S. Army Corps of Engineers Seattle harbor navigation project, operations and maintenance sampling and testing of Duwamish River sediments. Draft Report. 15 pp. + appendices.

Chan, S. M. Schiewe, K. Grams, A. Friedman, R. Bogar, U. Varanasi, W. Reichert, P.D. Plesha, S.J. Demuth, and D.W. Brown. 1986. East, West, and Duwamish Waterways Navigation Improvement Project: physical/chemical/biological analyses of sediments proposed for dredging. Final Report. Prepared by NOAA, National Marine Fisheries Service for U.S. Army Corps of Engineers, Seattle District. 106 pp.

Cohen, J. 1977. Statistical power analysis for the behavioral sciences. Academic Press, New York, NY.

DeWitt, T.H., G.R. Ditsworth, and R.C. Swartz. (1988). Effects of natural sediment features on survival of the phoxocephalid amphipod *Rhepoxynius abronius*. Mar. Environ. Res. 25:99-124.

Diachenko, G.W. 1981. Sorptive interactions of selected volatile hydrocarbons with humic acids from different environments. PhD Dissertation, University of Maryland.

DiToro, D.M. and L.M. Horzempa. 1982. Reversible and resistant components of PCB adsorption-desorption: isotherms. Environ. Sci. Technol. 16:594-602.

DiToro, D.M., J.S. Jeris, and D. Ciarcia. 1985. Diffusion and partitioning of hexachlorobiphenyl in sediments. Environ. Sci. Technol. 19:1169-1176.

Farrington, J.W. and J.M. Teal. 1982. Pollutant chemical measurements and biological availability: a lesson to be learned from fossil fuel hydrocarbons. Trans. Am. Geophys. Union. 53:86.

Fisher, J.B., R.L. Petty, and W. Lick. 1983. Release of polychlorinated biphenyls from contaminated lake sediments: flux and apparent diffusivities of four individual PCBs. Environ. Pollut. (Ser. B) 5:121-132.

Freeman, D.H. and L.S. Cheung. 1981. A gel partition model for organic desorption from a pond sediment. Science (Wash., DC) 214:790-792.

Hatcher, P.G., R. Rowan, and M.A. Mattingly. 1980. ^{1}H and ^{13}C NMR of marine humic acids. Org. Geochem. 2:77-85.

Hatcher, P.G., G.E. Maciel, and L.W. Dennis. 1981. Aliphatic structure of humic acids; a clue to their origin. Org. Geochem. 3:43-48.

Karickhoff, S.W. 1984. Organic pollutant sorption in aquatic systems. J. Hydraulic Engin. 110:707-735.

Karickhoff, S.W., and K.R. Morris. 1985a. Impact of tubificid oligochaetes on pollutant transport in bottom sediments. Environ. Sci. Technol. 19:51-56.

Karickhoff, S.W., and K.R. Morris. 1985b. Sorption dynamics of hydrophobic pollutants in sediment suspensions. Environ. Toxicol. Chem. 4:469-479.

Lie, U. 1968. A quantitative study of benthic infauna in Puget Sound, Washington, USA, in 1963-64. FiskDir. Skr., (Ser Havunders.). 14:229-556.

Long, E.R., and P.M. Chapman. 1985. A sediment quality triad: Measures of sediment contamination, toxicity and infaunal community composition in Puget Sound. Mar. Pollut. Bull. 16:405-415.

Mearns, A., R. Swartz, J. Cummins, D. Dinnell, P. Plesha, and P. Chapman. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod *Rhepoxynius abronius*. Mar. Environ. Res. 19:13-37.

Miller, R.G. 1981. Simultaneous statistical inference. Springer-Verlag, New York, NY. 299 pp.

Nichols, F.H. 1975. Dynamics and energetics of three deposit-feeding benthic invertebrate populations in Puget Sound, Washington. Ecol. Monogr. 45:57-82.

Nielsen, D. 1988. SEDQUAL Users Manual. Prepared by PTI Environmental Services for Tetra Tech/U.S. Environmental Protection Agency, Region X, Office of Puget Sound.

Osborn, J.G., D.E. Weitkamp, and T.H. Schadt. 1985. Alki wastewater treatment plant outfall improvements predesign study. Tech. Rep. No. 6.0, Marine Biology. Municipality of Metropolitan Seattle. 50 pp.

Pearson, T.H. and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanogr. Mar. Biol. Ann. Rev. 16:229-311.

Phillips, K., D. Jamison, J. Malek, B. Ross, C. Krueger, J. Thornton, and J. Krull. 1988. Evaluation procedures technical appendix: Sampling, testing, and test interpretation of dredged material proposed for unconfined, open-water disposal in central Puget Sound. Public Review draft report prepared by the Evaluations Procedures Work Group with assistance of Resource Planning Associates, PTI Environmental Services, Shapiro & Associates, and Tetra Tech, Inc. for the Puget Sound Dredged Disposal Analysis.

Prahl, F.G., and R. Carpenter. 1983. PAH-phase associations in Washington coastal sediment. Geochim. Cosmochim. Acta 47:1013-1023.

Reish, D.L., and J.L. Barnard. 1979. Amphipods (Arthopoda: Crustacea: Amphipoda). pp. 345-370. In: Pollution Ecology of Estuarine Invertebrates. C.W. Hart, Jr. and L.H. Fuller (eds).

Romberg, G.P., S.P. Pavlou, and E.A. Crecelius. 1984. Presence, distribution, and fate of toxicants in Puget Sound and Lake Washington. METRO Toxicant Program Report No. 6A. Toxicant Pretreatment Planning Study Technical Report C1. Municipality of Metropolitan Seattle, Seattle, WA. 231 pp.

Scheffe, H. 1959. The analysis of variance. John Wiley and Sons, New York, NY. 477 pp.

Sokal, R.R., and F.J. Rohlf. 1981. Biometry. 2nd Edition. W.H. Freeman and Co., San Francisco, CA. 859 pp.

Swartz, R.C., W.A. DeBen, J.K.P. Jones, J.O. Lamberson, and F.A. Cole. 1985. Phoxocephalid amphipod bioassay for marine sediment toxicity. pp. 284-307. In: Aquatic Toxicology and Hazard Assessment: Seventh Symposium. R.D. Cardwell, R. Purdy, and R.C. Bahner (eds). ASTM STP 854. American Society of Testing and Materials, Philadelphia, PA.

Tetra Tech. 1987. Recommended season for sampling subtidal benthic macroinvertebrates during Puget Sound monitoring. Draft Report. Prepared for U.S. Environmental Protection Agency, Seattle, WA. Tetra Tech, Inc., Bellevue, WA.

Trial, W., and J. Michaud. 1985. Alki wastewater treatment plant outfall improvements predesign study. Tech. Rep. No. 8.3, Water Quality. Municipality of Metropolitan Seattle. 89 pp.

U.S. Department of the Navy. 1985. Final environmental impact statement. Carrier Battle Group Puget Sound region ship homeporting project. Technical Appendix. Vol. 2. Prepared for U.S. Department of the Navy, Western Division, Naval Facilities Engineering Command, San Bruno, CA.

U.S. Environmental Protection Agency and U.S. Corps of Engineers Technical Committee on Criteria for Dredged and Fill Material. 1977. Ecological evaluation of proposed discharge of dredged material into ocean waters; implementation manual for Section 103 of Public Law 92-532 (Marine Protection, Research, and Sanctuaries Act of 1972). July 1977 (second printing April 1978). Environmental Effects Laboratory, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Word, J.Q., P.L. Striplin, K. Keeley, J. Ward, P. Sparks-McConkey, L. Bentler, S. Hulseman, K. Li, J. Schroeder, and K. Chew. 1984. Renton sewage treatment plant project: Seahurst baseline study, Vol. V, Section 6, Subtidal Benthic Ecology. FRI-UW-8413. University of Washington, Seattle, WA. 272 pp. + appendices.

GLOSSARY

Adverse Biological Effect

Any change in a biological system that results in injury or damage to an organism, population, or community (e.g., failure to develop properly, death, reduced population abundance).

Anomalous

Unusual observation or localized phenomena.

(1) Anomalous "no effect" - An observation of "no adverse biological effect" at a station when one was predicted by comparison of chemical data with sediment quality values. For example, **biological anomalies** can result from large variance in field or laboratory measurements leading to insufficient power to detect adverse effects. A similar apparent **chemical anomaly** could also result from predictions made by sediment quality values generated for one kind of sediment matrix to another matrix that reduces the bioavailability of contaminants.

(2) Anomalous "Effect" - An observation of "significant biological effects" at a station where AET were not exceeded. This kind of **biological anomaly** could result from a Type I error in the statistical test for biological effects. Apparent **chemical anomalies** could also result from interactive effects of chemicals or effects of chemicals that were not analyzed.

Amphipods

Small shrimp-like crustaceans (for example, sand fleas). Many live on the bottom, feed on algae and detritus, and serve as food for many marine species. Amphipods are used in laboratory bioassays to test the toxicity of sediments.

Amphipod Mortality

A measurement of amphipod death expressed as the percentage of individuals that die within 10 days upon exposure to test sediments in the laboratory. The standard protocol involves 5 replicate test chambers containing 20 amphipods each.

Apparent Effects Threshold

The sediment concentration of a contaminant above which statistically significant ($P \leq 0.05$) adverse biological effects (relative to appropriate reference conditions) would always be expected. An AET is calculated for each chemical and biological indicator.

Apparent Effects Threshold (cont.)

Defined AET An AET determined for a given chemical and given biological effect from a data set with at least one impacted station that exhibited a chemical concentration higher than the AET.

Preliminary AET An AET equal to the highest chemical concentration found in a particular data set (i.e., no impacted stations with a chemical concentration higher than the AET).

Highest AET The highest AET value (for a given chemical) among a set of AET for various biological indicators.

Lowest AET The lowest AET value (for a given chemical) among a set of AET for various biological indicators.

Benthic Infauna Macroinvertebrate organisms that live in or on the bottom of a body of water.

Benthic Infauna Abundance The population density of benthic infauna typically expressed as the number of organisms per unit area of bottom (e.g., number per m^2), as estimated from field sampling.

Bioassay A laboratory test used to evaluate the toxicity of a material (commonly sediments or wastewater) by measuring behavioral, physiological, or lethal responses of organisms.

Bulk Sediment Analysis Chemical analyses performed on an entire sediment sample, without separating water from the solid material in a sample.

Cause-Effect Relationship The quantitative interdependence of a biological effect variable and an environmental variable such as the concentration of a certain chemical, organic carbon, or sediment grain size. Cause-effect relationships cannot be proven using field-based approaches such as AET.

Chemical Interactive Effects

Combined action of two or more chemicals to produce additive, synergistic, or antagonistic effects on an organism, a population, or a biological community. An additive effect results when the combined effects can be accurately predicted by the sum of the individual effects of the single chemicals. A synergistic effect results when the combined effects are greater than the sum of the individual chemical effects. An antagonistic effect results when the combined effects are less than the sum of the individual chemical effects.

Contaminant

A chemical or biological substance in a form or in a quantity that can harm aquatic organisms, consumers of aquatic organisms, or users of the aquatic environment.

Contaminated Sediment

Technical Definition: A sediment that contains measurable levels of contaminants.

Management Definition: A sediment that contains sufficient quantities of contaminants to result in adverse environmental effects and thus require restrictions for dredging of dredged material, or consideration of remedial action.

Conventional Pollutants

Sediment variables and characteristics that have been routinely measured in assessing sediment quality. These include sulfides, total organic carbon, total organic nitrogen, total volatile solids, and grain size distribution (e.g., percent coarse and fine-grained material).

Detection Limit

The lowest concentration at which the presence of a particular chemical can be reliably established by a specified analytical method. Detection limits used in this report may in some cases be lower than strictly defined method detection limits (i.e., the concentration of a chemical above which there is 99 percent statistical confidence that the detected signal exceeds background noise levels).

Dredged Material

Sediments excavated from the bottom of a waterway or water body.

Efficiency

A measure of reliability for any sediment quality value. The proportion of all stations predicted to have adverse biological effects that are actually impacted.

Empirical

Depending on or based on observations in the laboratory or the field. Also, subject to verification by observation or experiment.

Equilibrium Partitioning Approach

A theory of the distribution of chemicals in various phases (e.g., dissolved, particulate, gaseous) based on thermodynamic principles. The equilibrium partitioning approach has been used to predict the concentrations of chemicals in bulk sediments from corresponding concentrations in interstitial (pore) water, or conversely, to predict the concentrations of chemicals in interstitial water from corresponding concentrations in bulk sediment.

Hydrocarbon

An organic compound composed of carbon and hydrogen. Petroleum and its derived compounds consist primarily of hydrocarbons.

Impacted

Changed from a natural state as a result of human activities. Typically, the term "impacted" is applied to describe the status of biological systems adversely affected by human activities.

Intertidal Area

The area between high and low tide levels. The alternate wetting and drying of this area makes it a transition between land and water organisms and creates special environmental conditions.

Maximum Level

A sediment quality value based on the highest AET for a range of biological indicators and used in PSDDA to evaluate when biological testing may be unnecessary to determine that a contaminated sediment is unsuitable for unconfined, open-water disposal.

Metals

Naturally occurring elements. Certain metals, such as mercury, lead, nickel, zinc, and cadmium can be or environmental concern when they are released to the environment in unnatural amounts by man's activities.

Microtox

A laboratory test using luminescent bacteria and measuring light production, used to assess toxicity of saline or organic solvent extracts of sediments.

Noncontradictory	Consistent with available evidence (i.e., for AET, correct prediction of all nonimpacted stations in the dataset used to generate AET is part of the noncontradictory assumption in the AET approach).
Nonimpacted	Not affected adversely by human activities.
Organic Carbon	Carbon in a form related to or derived from living things (excluding calcium carbonate).
Oyster Larvae Abnormality	A measurement of developmental defects in oyster expressed as the percentage of individuals that fail to develop properly within 48 hours upon exposure to test sediments in the laboratory. The standard protocol involves 5 replicate test chambers, each containing embryos at a density of 20-40 per milliliter.
Pesticide	A general term used to describe any substance, usually chemical, used to destroy or control organisms (pests). Pesticides include herbicides, insecticides, algicide, and fungicides. Many of these substances are manufactured and are not naturally found in the environment. Others are natural toxins that are extracted from plants and animals.
Polycheate	A marine worm of the Phylum Annelida, Class Polychaeta.
Polychlorinated Biphenyls	A group of manmade organic chemicals, including 209 different but closely related compounds (congeners) made up of carbon, hydrogen and chlorine. If released to the environment, they persist for long periods of time and can concentrate in food chains. The manufacture and use of PCBs are regulated by EPA under the Toxic Substances Control Act.
Power Analysis	A mathematical technique to determine the capacity of a statistical test to detect true differences between treatments (e.g., between amphipod mortality measured for a test sediment vs. that of a reference-area sediment). Statistical power is expressed as the probability of detecting a true difference of a specified magnitude.

Priority Pollutants

Substances listed by EPA under the Clean Water Act as toxic and having priority for regulatory controls. The list includes toxic metals, inorganic contaminants such as cyanide and arsenic, and a broad range of both natural and artificial organic compounds. The list of priority pollutants includes some substances that are not of concern in Puget Sound, and also does not include all known harmful compounds.

Puget Sound Water Quality Authority

An agency created by the Washington State legislature in 1985 and tasked with developing a management plan to protect and enhance the water quality of Puget Sound. The Authority adopted its first plan in January 1987.

Reliability

A general quality resulting from a combination of both high sensitivity and high efficiency of sediment quality values.

Screening Level

A sediment quality value based partly on AET modified with a safety factor and used in PSDDA to evaluate when biological testing should be conducted to determine the suitability of the sediment for unconfined, open-water disposal.

Sediment

Material suspended in or settling to the bottom of a liquid, such as the sand and mud that make up much of the shorelines and bottom of Puget Sound. Sediment in Puget Sound comes from natural sources (e.g., erosion of soils and weathering of rock) or anthropogenic sources (e.g., forest or agricultural practices or construction activities). Certain contaminants tend to collect on and adhere to sediment particles. The sediment of some areas around Puget Sound contain elevated levels of contaminants relative to reference areas that are located away from major contaminant sources.

Sediment Quality Value

A "benchmark" number for an environmentally acceptable concentration of a given chemical in sediments. A chemical concentration that is expected to be below the level that would consistently lead to adverse biological effects for a wide variety of species, habitats, and sediment types.

SEDQUAL	Sediment quality data base and menu-driven program developed for Puget Sound Estuary Program
Sensitivity	A measure of reliability for any sediment quality value. The proportion of all stations exhibiting adverse biological effects that are correctly predicted by a sediment quality value.
Statistically Significant	A quantitative determination of the statistical degree to which multiple measurements of the same variable can be shown to be different, given the variability of the measurements.
Subtidal	Refers to the marine environment below low tide.
Synoptic	(or **Matched Data**) Data for different variables that are collected during a single sampling event at a single station (e.g., chemical data on sediments at a station that were also evaluated for benthic infauna abundances and sediment toxicity to amphipods).
Toxic Substances and Toxicants	Chemical substances that are poisonous, carcinogenic, or otherwise harmful to life if found in sufficient concentration.
Volatile Solids	The material in a sediment sample that vaporizes or evaporates at a given high temperature (550° C).

ABBREVIATIONS

AET	Apparent Effects Threshold
Authority	Puget Sound Water Quality Authority
Corps	U.S. Army Corps of Engineers
DNR	Washington Department of Natural Resources
Ecology	Washington Department of Ecology
EPA	U.S. Environmental Protection Agency
ML	Maximum Level (defined in the PSDDA program)
PAH	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyls
PSDDA	Puget Sound Dredged Disposal Analysis
PSEP	Puget Sound Estuary Program
SEDQUAL	SEDiment QUALity database
SL	Screening Level (defined in the PSDDA program)

APPENDIX A

CONCEPT OF THE AET APPROACH

CONTENTS

FIGURES

CONCEPT OF THE AET APPROACH

An AET is defined as the sediment concentration of a given chemical above which statistically significant ($P \leq 0.05$) biological effects (e.g., depressions in the abundances of indigenous benthic infauna) are always expected. If any chemical exceeds its AET for a particular biological indicator, an adverse biological effect is predicted for that indicator. If all chemical concentrations are below their AET for a particular biological indicator then no adverse effect is predicted.

In this appendix, AET generation is described and the AET concept is discussed as it relates to the interpretation of chemical and biological data in field-collected sediments. AET generation is a conceptually simple process, and incorporates the complexity of biological-chemical interrelationships in the environment without relying upon *a priori* assumptions as to the mechanistic nature of these interrelationships. The concept of the AET is presented in this section with little reference to specific chemicals, specific biological tests, or specific chemical normalizations, because the approach is not inherently limited to specific subsets of these variables.

DESCRIPTION OF THE AET APPROACH

The focus of the AET approach is to identify concentrations of contaminants that are associated exclusively with sediments exhibiting statistically significant biological effects relative to reference sediments. The calculation of AET for each chemical and biological indicator is straightforward:

1. **Collect "matched" chemical and biological effects data**--Conduct chemical and biological effects testing on subsamples of the same field sample (to avoid unaccountable losses of benthic organisms, benthic infaunal and chemical analyses are conducted on separate samples collected concurrently)

2. **Identify "impacted" and "nonimpacted" stations**--Statistically test the significance of adverse biological effects relative to suitable reference conditions for each sediment sample and biological indicator; suitable reference conditions are established by sediments containing very low or undetectable concentrations of any toxic chemicals

3. **Identify AET using only "nonimpacted" stations**--For each chemical, the AET can be identified for a given biological indicator as the highest detected concentration among sediment samples that do not exhibit statistically significant effects (if the chemical is undetected in all nonimpacted samples, no AET can be established for that chemical and biological indicator)

4. **Check for preliminary AET**--Verify that statistically significant biological effects are observed at a chemical concentration higher than the AET; otherwise the AET is only a preliminary minimum estimate (or may not exist).

5. Repeat Steps 1-4 for each biological indicator.

A pictorial representation of the AET approach for two example chemicals is presented in Figure A-1 based on results for a toxicity bioassay. Two subpopulations of all sediments analyzed for chemistry and subjected to a bioassay are represented by bars in the figure, and include:

- Sediments that did not exhibit statistically significant ($P>0.05$) toxicity relative to reference conditions ("nonimpacted" stations)
- Sediments that exhibited statistically significant ($P\leq0.05$) toxicity in bioassays relative to reference conditions ("impacted" stations).

The horizontal axis in each figure represents sedimentary concentrations of chemicals (lead or 4-methylphenol) on a log scale. Dry weight normalized data are presented in Figure A-1, although the AET approach is not limited to any particular normalization. For the toxicity bioassay under consideration, the AET for lead is the highest lead concentration corresponding to sediments that did not exhibit significant toxicity (the top bar for lead in Figure A-1). Above this lead AET, significant toxicity was always observed in the data set. The AET for 4-methylphenol was determined analogously.

INTERPRETATION OF AET

An AET corresponds to the sediment concentration of a chemical above which all samples for a particular biological indicator were observed to have adverse effects. Thus, the AET is based on non-contradictory evidence of biological effects. Data are treated in this manner to reduce the weight given to samples in which factors other than the contaminant examined (e.g., other contaminants, environmental variables) may be responsible for the biological effect.

Relationships Among Chemical-Specific AET

Using Figure A-1 as an example, sediment from Station SP-14 exhibited severe toxicity, potentially related to a greatly elevated level of 4-methylphenol (7,400 times reference levels). The same sediment from Station SP-14 contained a relatively low concentration of lead that was well below the AET for lead (Figure A-1). Despite the toxic effects associated with the sample, sediments from many other stations with higher lead concentrations than SP-14 exhibited no statistically significant biological effects. These results were interpreted to suggest that the effects at Station SP-14 were potentially associated with 4-methylphenol (or a substance with a similar environmental distribution) but were less likely to be associated with lead. A converse argument can be made for lead and 4-methylphenol in sediments from Station RS-18. In this manner, the AET approach helps to identify measured chemicals that are potentially associated with observed effects at each biologically impacted site and eliminates from consideration chemicals that are far less likely to be associated with effects (i.e., the latter chemicals have been observed at higher concentrations at other sites without associated biological effects). Based on the results for lead and 4-methylphenol, effects at 4 of the 28 impacted sites shown in the figures may be associated with elevated concentrations of 4-methylphenol, and effects at 7 other sites may be associated with elevated concentrations of lead (or similarly distributed contaminants).

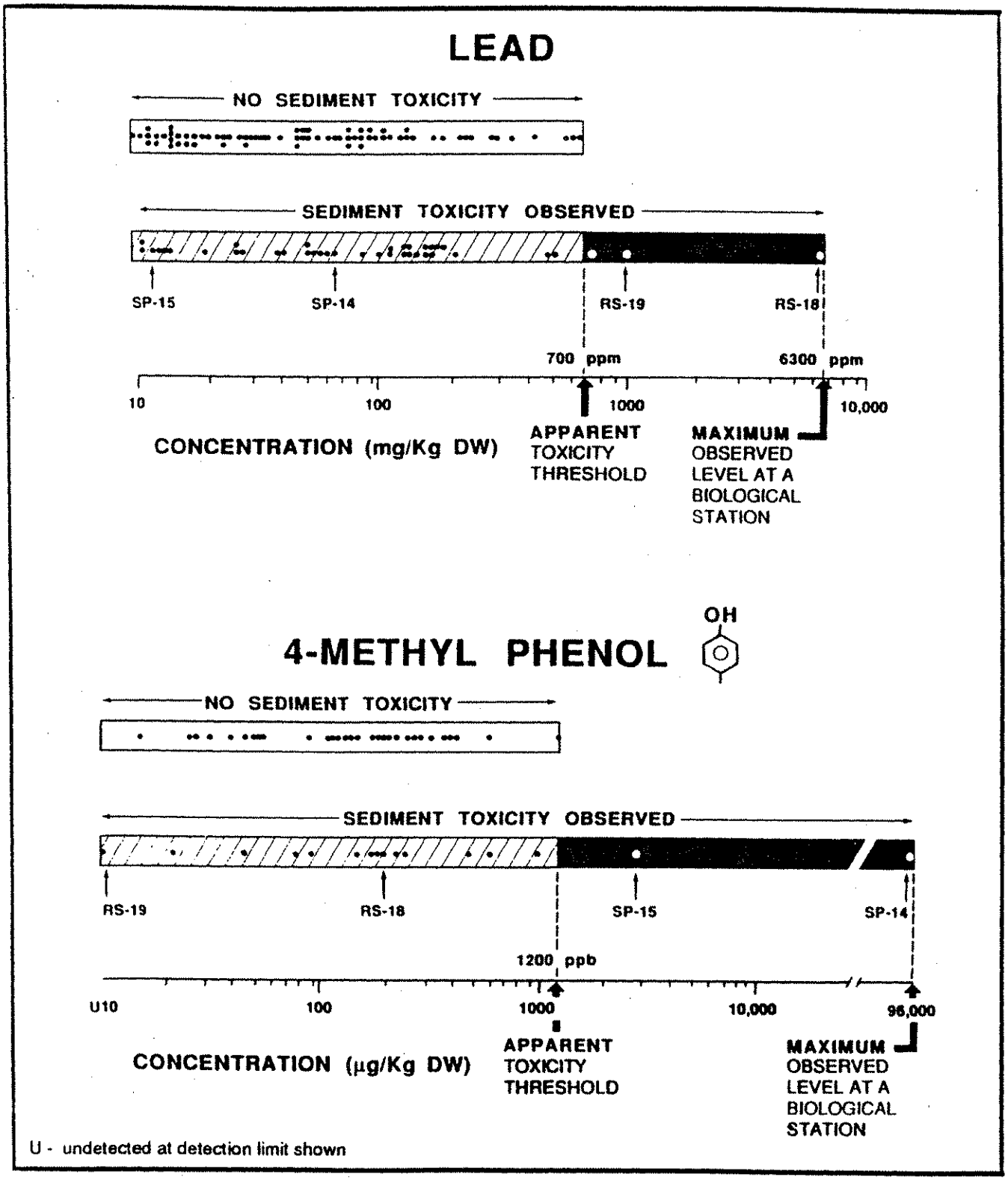


Figure A-1. The AET approach applied to sediments tested for lead and 4-methyl phenol concentrations and toxicity response during bioassays.

These results illustrate that the occurrence of biologically impacted stations at concentrations below the AET of a single chemical does not imply that AET in general are not protective against biological effects, only that single chemicals may not account for all stations with biological effects. By developing AET for multiple chemicals, a high percentage of all stations with biological effects are accounted for with the AET approach (reliability results are presented in Section 3 of the main report).

AET can be expected to be more predictive when developed from a large, diverse database with wide ranges of chemical concentrations and a wide diversity of measured chemicals. Data sets that have large concentration gaps between stations and/or that do not cover a wide range of concentrations must be scrutinized carefully (e.g., to discern whether chemical concentrations in the data set exceed reference concentrations) before generation of AET is appropriate.

Dose-Response Relationships and AET

The AET concept is consistent with empirical observations in the laboratory of dose response relationships between increasing concentrations of individual toxic chemicals and increasing biological effects. A simple hypothetical example of such single-chemical relationships is shown for chemicals X and Y in Figure A-2. In the example, data are shown for laboratory exposures of a test organism to sediment containing only increasing concentrations of chemical X, and independently, for exposures to sediment containing only increasing concentrations of chemical Y. The magnitude of toxic response in the example differs for the two chemicals, and occurs over two different concentration ranges. It is assumed that at some level of response, for example >25 percent, the two different responses can be distinguished from reference conditions (i.e., responses resulting from exposure to sediments containing very low or undetectable concentrations of any toxic chemicals).

These single-chemical relationships cannot be proven in the field because organisms are exposed to complex mixtures of chemicals in environmental samples. In addition, unrelated discharges from different sources can result in uncorrelated distributions of chemicals in environmental samples. To demonstrate the potential effects of these distributions, response data are shown in Figure A-3 for a random association of chemicals X and Y using the same concentration data as in Figure A-2. The data have been plotted according to increasing concentrations of chemical X and the same dose response relationship observed independently for the two chemicals in the laboratory has been assumed. The contributions of chemicals X and Y to the toxic response shown for these simple mixtures is intended only for illustration purposes to enable direct comparison to the relationships shown in Figure A-2, but is analogous to an additive toxic response. Other interactive effects are not considered in this example.

In Figure A-3, a significant response relative to reference conditions would result whenever elevated concentrations of either chemical X or chemical Y occurred in a sample. Because of the random association of Y with X in these samples, the significant responses would appear to occur randomly over the lower concentration range of chemical X. The classification of the responses shown in Figure A-3 into significant and nonsignificant groups (i.e., >25 percent response for either chemical) results in generation of Figure A-4.

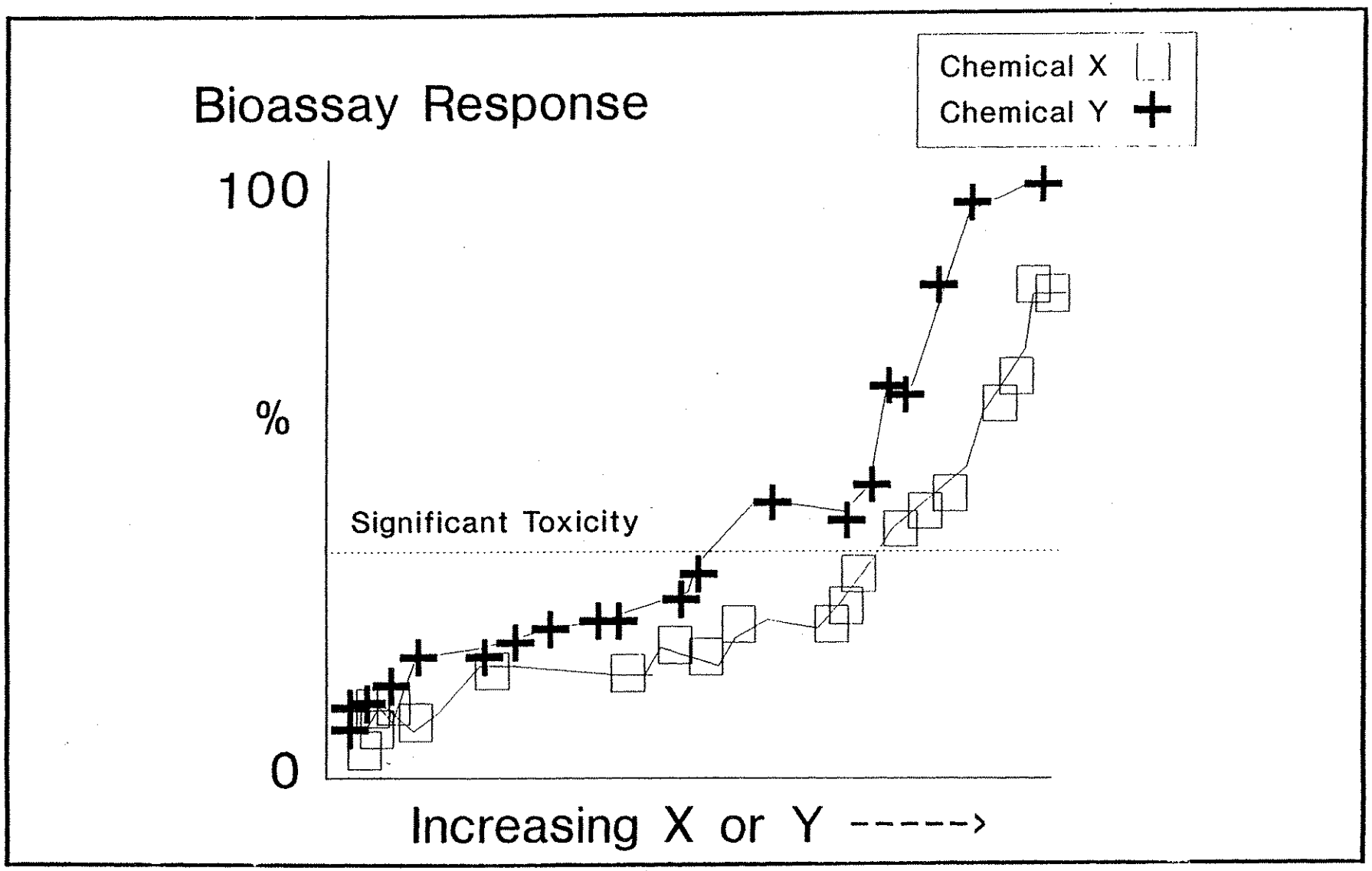


Figure A-2. Hypothetical example of dose-response relationship resulting from laboratory exposure to single chemicals X and Y.

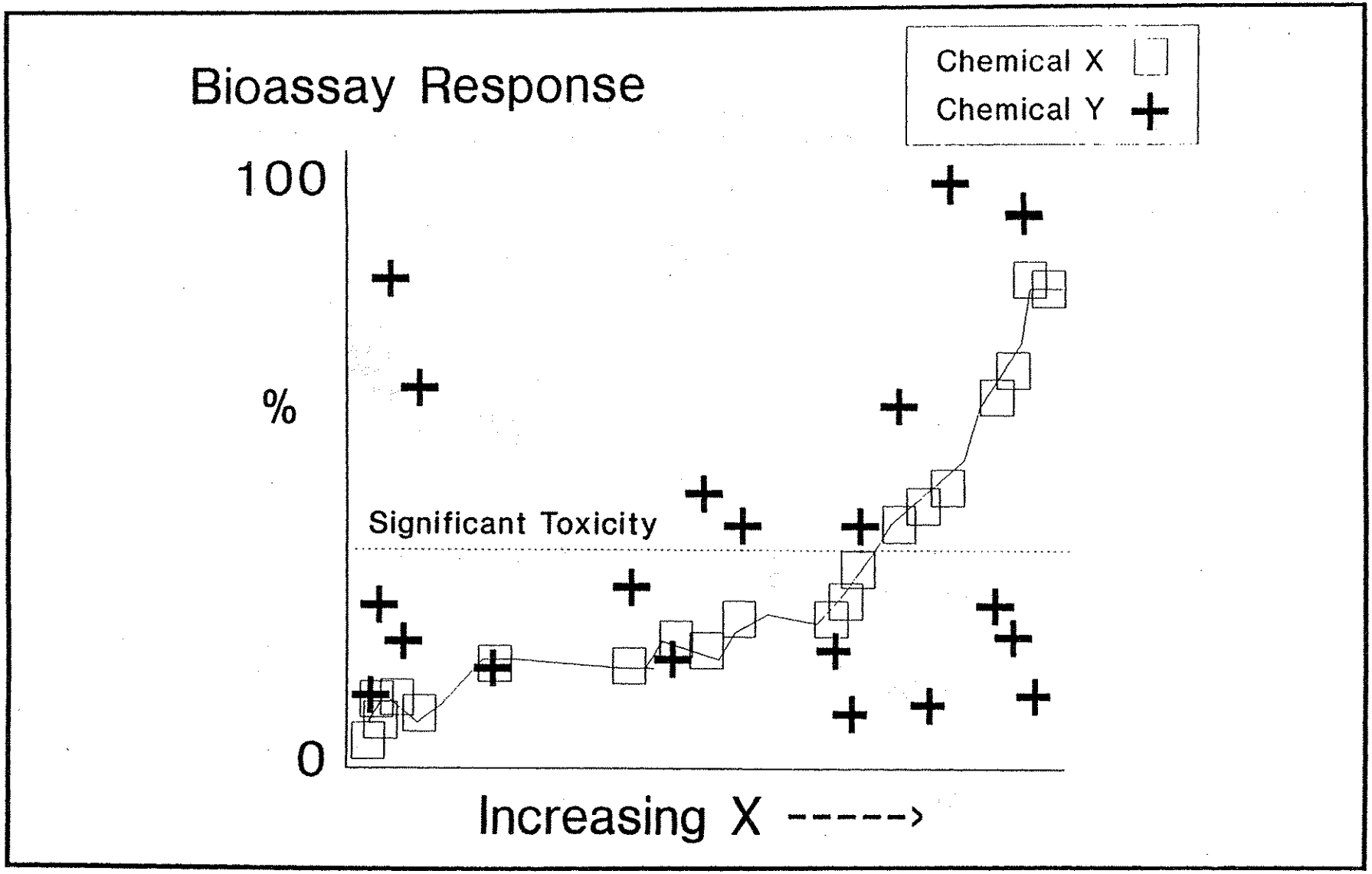


Figure A-3. Hypothetical example of toxic response resulting from exposure to environmental samples of sediment contaminated with chemicals X and Y.

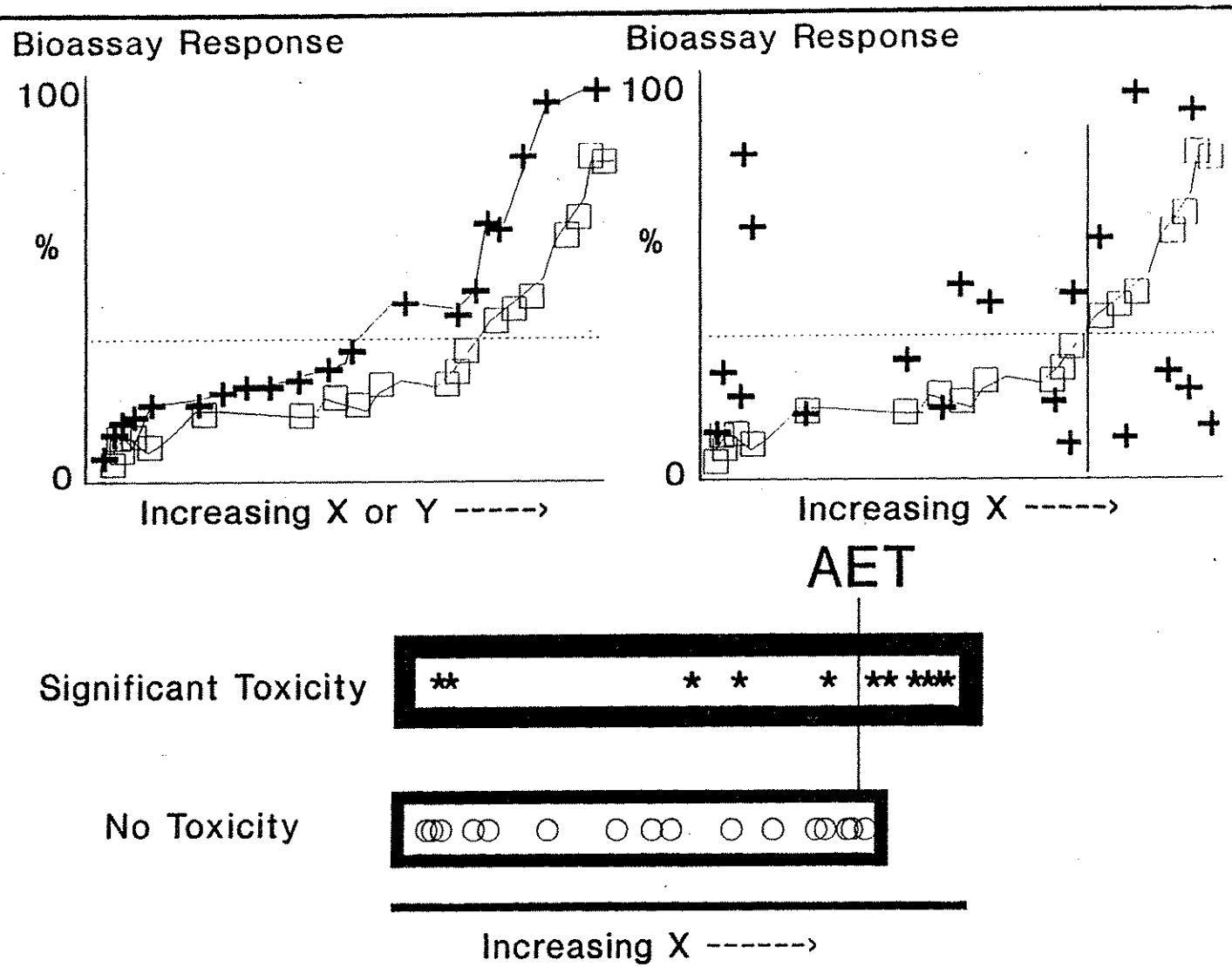


Figure A-4. Hypothetical example of AET calculation for chemical X based on classification of significant and nonsignificant responses for environmental samples contaminated with both chemicals X and Y. [Previous Figures A-2 and A-3 are also shown for comparison; dashed line indicates level of significant toxicity]

Figure A-4 represents how the actual environmental results would appear when ranked according to concentration of chemical X using these same data. Below the AET for chemical X, significant toxicity is produced by elevated concentrations of chemical Y, which is randomly associated with the distribution of chemical X. Above the AET for chemical X, significant toxicity is always produced by elevated concentrations of chemical X although in some samples, elevated concentrations of chemical Y also contribute to the overall toxicity. The AET for chemical X corresponds conceptually, in this simple example, to the concentration in Figure A-2 at which a significant difference in response was observed in the laboratory for chemical X.

In environmental samples that contain complex mixtures of chemicals, a monotonic dose-response relationship such as in this simple two-chemical example may not always apply. For example, a consistently increasing biological response may not always occur at increasing concentrations of a chemical above its AET. Such observations could indicate that the AET is coincidental (i.e., that the observed toxicity in some or all samples above the AET is unrelated to the presence of that chemical), or that changing environmental factors in samples exceeding an AET obscure a monotonic dose-response relationship. Such factors are discussed in the following section.

Influence of Environmental Factors on AET Interpretation

Although the AET concept is simple, the generation of AET values based on environmental data incorporates many complex biological-chemical interrelationships. For example, the AET approach incorporates the net effects of the following factors that may be important in field-collected sediments:

- Interactive effects of chemicals (e.g., synergism, antagonism, and additivity)
- Unmeasured chemicals and other unmeasured, potentially adverse variables
- Matrix effects and bioavailability (i.e., phase associations between contaminants and sediments that affect bioavailability of the contaminants, such as the incorporation of PAH in soot particles).

The AET approach cannot distinguish and quantify the contributions of interactive effects, unmeasured chemicals, or matrix effects in environmental samples, but AET values may be influenced by these factors. To the extent that the samples used to generate AET are representative of samples for which AET are used to predict effects, the above environmental factors may not detract from the predictive reliability of AET. Alternatively, the infrequent occurrence of the above environmental factors in a data set used to generate AET could detract from the predictive reliability of those AET values. If confounding environmental factors render the AET approach unreliable, this should be evident from validation tests in which biological effects are predicted in environmental samples. Tests of AET values generated from Puget Sound data (see Section 3 of the main report) indicate that the approach is relatively reliable in predicting biological effects despite the potential uncertainties of confounding environmental factors.

Although the above environmental factors can influence the *generation* of field-based sediment quality values such as AET, they also may influence the *application* of all sediment quality value approaches for the prediction of adverse biological effects. For example, sediment quality values based on laboratory sediment bioassays spiked with

single chemicals would not be susceptible to the effects of the environmental factors listed above; however, in applying such values to field-collected samples, predictions of biological effects could be less successful to the extent that interactive effects, unmeasured chemicals, and matrix effects occur in the environment.

The nature of the relationships between AET values and confounding environmental factors is discussed in the remainder of this section.

Interactive Effects and AET -- AET uncertainty is increased by the possibility of interactive effects; the increase in uncertainty is expected to be less pronounced when large data sets collected from diverse areas are used to generate AET. Additivity and synergism can produce a comparatively low AET for a given chemical by causing impacts at concentrations that would not cause impacts in the absence of these interactive effects. This would effectively reduce the pool of nonimpacted stations used to generate AET. This effect should be reduced if a diverse database is used such that chemicals occur over a wide range of concentrations at stations where additivity and synergism are not operative. For chemicals that covary regularly in the environment (e.g., fluoranthene and pyrene), even a large, diverse database will not reduce the effects of additivity and/or synergism on AET generation. The resulting AET values for such chemicals may be reliable in predicting biological effects in environmental samples although not representative of the toxicities of the chemicals acting independently.

Antagonism will produce comparatively high AET values if (and only if) the AET is established at a station where antagonism occurs. A large, diverse database could not rectify this elevation of AET if the station at which antagonism occurred was the nonimpacted station with the highest concentration (i.e., the station setting an AET). An AET set by a station at which antagonism occurred would not be representative of the toxicity of the chemical acting independently. Hence, if antagonism did not occur widely, such antagonistic effects would cause the AET to be less sensitive in predicting adverse effects in the environment.

Empirical approaches such as the AET do not provide a means for characterizing interactive effects. Only laboratory-spiked sediment bioassays offer a systematic and reliable method for identifying and quantifying additivity, synergism, and antagonism. Unfortunately, a great deal of research effort would be required to test the range of chemicals potentially occurring in the environment (both individually and in combination) and a sufficiently wide range of organisms to establish criteria. In addition, the applicability of bioassays conducted with laboratory-spiked sediments to environmentally-contaminated sediments requires further testing.

Unmeasured Chemicals and AET--Another source of uncertainty for AET and other field-based approaches is the possibility of effects being caused by unmeasured, covarying chemicals. Such chemicals would not be expected to substantially decrease the ability of AET to predict biologically impacted stations (excluding interactive effects discussed above). If an unmeasured chemical (or group of chemicals) varies consistently in the environment with a measured chemical, then the AET established for the measured contaminant will indirectly apply to, or result in the management of, the unmeasured contaminant. In such cases, a measured contaminant would act as a surrogate for an unmeasured contaminant (or group of unmeasured contaminants). Because all potential contaminants cannot be measured routinely, management strategies must rely to some extent on "surrogate" chemicals.

If an unmeasured toxic chemical (or group of chemicals) does not always covary with a measured chemical (e.g., if a certain industry releases an unusual mixture of contaminants), the effect should be mitigated if a sufficiently large and diverse data set is used to establish AET. Use of a large data set comprising samples from a variety of areas with wide-ranging chemical concentrations would decrease the likelihood that an unrealistically low AET would be set. Because AET are set by the highest concentration of a given chemical in samples without observed biological effects, AET will not be affected by less contaminated samples in which unmeasured contaminants cause biological effects.

If an unmeasured toxic chemical does not covary with any of the measured chemicals, it is unlikely that the AET (or any other chemical-specific approach) could predict impacts at stations where the chemical is inducing toxic effects. The frequency of occurrence of stations with biological effects but no chemicals exceeding AET is the subject of validation tests (see Section 3 of the main report).

Matrix Effects and Bioavailability--Geochemical associations of contaminants with sediments that reduce bioavailability of those contaminants would affect AET analogously to antagonistic effects (i.e., they would increase AET relative to sediments in which this factor was not operative). Sediment matrices observed in Puget Sound that may reduce bioavailability of certain contaminants include slag material (containing high concentrations of various metals and metalloids, such as copper and arsenic) and coal or soot (which may contain high concentrations of largely unavailable PAH, as opposed to oil or creosote, in which PAH would be expected to be far more bioavailable; e.g., Farrington and Teal 1982). Many kinds of matrices may occur in the environment and a large proportion may be difficult to classify based upon appearance or routinely measured sediment variables. Hence, the use of matrix-specific data sets to generate AET, although desirable, would be difficult to implement.

To address this concern from a technical perspective (i.e., representativeness of data used in AET generation), the AET database could be screened for sediment with chemical concentrations that are anomalously high relative to those in other nonimpacted sediments from different geographic areas. From a management perspective, this guideline would generate more protective (sensitive) sediment quality standards that may also be less efficient in only identifying problem sediments. These sediments would be considered nonrepresentative and not used in AET generation unless and until additional data could substantiate that they are representative. Such data treatment methods are discussed in Volume I, Appendix C of this report.

APPENDIX B

SUMMARY OF DATA SOURCES

FIGURES

TABLES

TABLE B-1. SUMMARY OF SURVEYS USED TO EVALUATE PUGET SOUND AET

Survey	Survey Name	Sampling Date	Responsible Agency
ALKI	1984 Alki Pt. Survey	5/84	METRO
CBBLAIR	Commencement Bay Blair Waterway Dredging Survey	6/84	Port of Tacoma
CBMSQS	Commencement Bay RI Main Sediment Quality Survey	3/84	Washington Dept. of Ecology, U.S. EPA, Region 10
DUWRIV1	Duwamish River I Dredged Material Characterization Survey	4/85	U.S. Army Corps of Engineers
DUWRIV2	Duwamish River II Dredged Material Characterization Survey	9/85	U.S. Army Corps of Engineers
EBCHEM	Elliott Bay Toxics Action Plan	9/85	U.S. EPA, Region 10
EHCHEM	Eagle Harbor Preliminary Survey	6/85	Washington Dept. of Ecology
EIGHTBAY	Puget Sound Eight-Bay Reconnaissance Survey	5/84	U.S. EPA, Region 10
EVCHEM	Everett Harbor Toxics Action Plan	9/86	U.S. EPA, Region 10
EVERETT1	Navy home-port EIS project	/85	U.S. Navy
TPPS3AB	Toxics Pretreatment Planning Study, Phase III A & B	3/82 7/82	METRO

TABLE B-2. BIOLOGICAL EFFECTS STATIONS USED TO EVALUATE 1986 AND 1988 PUGET SOUND AET.

The biological effects listed are:

AMPT Amphipod toxicity AET.
BENA Benthic infauna abundance AET.
MICB Microtox bioassay AET.
OYST Oyster larvae toxicity AET.

Survey	Station	Date	Sample	AMPT	BENA	MICB	OYST
ALKI	AP-01	05/25/84	AP-01		NO HITa		
ALKI	AP-02	05/25/84	AP-02		NO HIT		
ALKI	AP-03	05/25/84	AP-03		NO HIT		
ALKI	AP-04	05/25/84	AP-04		HIT		
ALKI	AP-05	05/25/84	AP-05		NO HIT		
ALKI	AP-06	05/25/84	AP-06		NO HIT		
ALKI	AP-07	05/25/84	AP-07		NO HIT		
ALKI	PW-01	05/26/84	PW-01		NO HIT		
ALKI	PW-02	05/26/84	PW-02		NO HIT		
ALKI	PW-03	05/26/84	PW-03		NO HIT		
ALKI	PW-04	05/26/84	PW-04		NO HIT		

*Benthic infauna abundance was the only biological effect tested in the Alki survey.

CBBLAIR	B03	06/01/84	B03	NO HIT	NO HIT		NO HIT
CBBLAIR	B04	06/01/84	B04	NO HIT	NO HIT		NO HIT
CBBLAIR	B09	06/01/84	B09	NO HITa	NO HIT		NO HIT
CBBLAIR	B10	06/01/84	B10	NO HITa	NO HIT		HIT
CBBLAIR	B12	06/01/84	B12	NO HITa	NO HIT		NO HIT
CBBLAIR	B15	06/01/84	B15	HIT	NO HIT		NO HIT

*Benthic infauna abundance, and the amphipod and oyster larvae toxicity tests were the only biological effects tested in the CBBLAIR survey.

CBMSQS	BL-11	01/01/84	BL-11	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	BL-13	01/01/84	BL-13	NO HITa	NO HIT	HIT	NO HIT
CBMSQS	BL-21	01/01/84	BL-21	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	BL-25	01/01/84	BL-25	HIT	NO HIT	NO HIT	NO HIT
CBMSQS	BL-28	01/01/84	BL-28	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	BL-31	01/01/84	BL-31	NO HITa	NO HIT	HIT	NO HIT
CBMSQS	CI-11	01/01/84	CI-11	HIT	HIT	HIT	HIT
CBMSQS	CI-13	01/01/84	CI-13	NO HITa	HIT	HIT	HIT
CBMSQS	CI-16	01/01/84	CI-16	b	HIT	HIT	HIT
CBMSQS	CI-17	01/01/84	CI-17	NO HITa	HIT	HIT	NO HIT
CBMSQS	CI-20	01/01/84	CI-20	HIT	NO HITa	NO HIT	HIT
CBMSQS	CI-22	01/01/84	CI-22	NO HITa	NO HIT	NO HIT	NO HIT
CBMSQS	CR-11	01/01/84	CR-11	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	CR-12	01/01/84	CR-12	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	CR-13	01/01/84	CR-13	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	CR-14	01/01/84	CR-14	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	HY-12	01/01/84	HY-12	b	NO HIT	HIT	HIT
CBMSQS	HY-14	01/01/84	HY-14	NO HIT	HIT	HIT	NO HIT
CBMSQS	HY-17	01/01/84	HY-17	NO HITa	HIT	HIT	HIT
CBMSQS	HY-22	01/01/84	HY-22	HIT	HIT	HIT	HIT

CBMSQS	HY-23	01/01/84	HY-23	HIT	HIT	HIT	HIT
CBMSQS	HY-24	01/01/84	HY-24	NO HITa	HIT	HIT	NO HIT
CBMSQS	HY-28	01/01/84	HY-28	NO HIT	HIT	NO HIT	NO HIT
CBMSQS	HY-32	01/01/84	HY-32	NO HITa	HIT	NO HIT	NO HIT
CBMSQS	HY-37	01/01/84	HY-37	NO HITa	HIT	HIT	NO HIT
CBMSQS	HY-42	01/01/84	HY-42	HIT	HIT	HIT	NO HIT
CBMSQS	HY-43	01/01/84	HY-43	NO HITa	HIT	HIT	NO HIT
CBMSQS	HY-44	01/01/84	HY-44	NO HITa	NO HIT	NO HIT	NO HIT
CBMSQS	HY-47	01/01/84	HY-47	HIT	HIT	HIT	HIT
CBMSQS	HY-50	01/01/84	HY-50	NO HITa	NO HIT	HIT	NO HIT
CBMSQS	MD-12	01/01/84	MD-12	NO HIT	HIT	NO HIT	NO HIT
CBMSQS	MI-11	01/01/84	MI-11	HIT	NO HIT	NO HIT	NO HIT
CBMSQS	MI-13	01/01/84	MI-13	NO HIT	NO HIT	HIT	NO HIT
CBMSQS	MI-15	01/01/84	MI-15	HIT	NO HIT	HIT	NO HIT
CBMSQS	RS-12	01/01/84	RS-12	NO HIT	NO HIT	NO HIT	NO HIT
CBMSQS	RS-13	01/01/84	RS-13	HIT	NO HIT	NO HIT	HIT
CBMSQS	RS-14	01/01/84	RS-14	NO HITa	NO HITa	NO HIT	NO HIT
CBMSQS	RS-18	01/01/84	RS-18	HIT	HIT	HIT	HIT
CBMSQS	RS-19	01/01/84	RS-19	HIT	HIT	HIT	HIT
CBMSQS	RS-20	01/01/84	RS-20	NO HIT	HIT	HIT	NO HIT
CBMSQS	RS-22	01/01/84	RS-22	NO HIT		NO HIT	NO HIT
CBMSQS	RS-24	01/01/84	RS-24	HIT		NO HIT	NO HIT
CBMSQS	SI-11	01/01/84	SI-11	NO HITa	HIT	HIT	NO HIT
CBMSQS	SI-12	01/01/84	SI-12	HIT	HIT	HIT	NO HIT
CBMSQS	SI-15	01/01/84	SI-15	HIT	NO HITa	NO HIT	NO HIT
CBMSQS	SP-11	01/01/84	SP-11	NO HIT	HIT	HIT	NO HIT
CBMSQS	SP-12	01/01/84	SP-12	NO HITa	NO HIT	HIT	HIT
CBMSQS	SP-14	01/01/84	SP-14	HIT	HIT	HIT	HIT
CBMSQS	SP-15	01/01/84	SP-15	HIT	HIT	HIT	HIT
CBMSQS	SP-16	01/01/84	SP-16	HIT	HIT	HIT	HIT
DUWRIV1	DR-01	04/15/85	DR-01	NO HIT			
DUWRIV1	DR-02	04/15/85	DR-02	NO HIT			
DUWRIV1	DR-03	04/15/85	DR-03	NO HIT			
DUWRIV1	DR-04	04/15/85	DR-04	NO HIT			
DUWRIV1	DR-05	04/15/85	DR-05	NO HIT			
DUWRIV1	DR-06	04/15/85	DR-06	NO HIT			
DUWRIV1	DR-07	04/15/85	DR-07	HIT			
DUWRIV1	DR-08	04/15/85	DR-08	HIT			
DUWRIV1	SQ-09	04/19/85	SQ-09	NO HIT			
DUWRIV2	DR-10	07/01/85	CA1	HIT			
DUWRIV2	DR-11	07/01/85	CA2	HIT			
DUWRIV2	DR-12	07/01/85	CA3	NO HIT			
DUWRIV2	DR-13	07/01/85	CB1	NO HIT			
DUWRIV2	DR-14	07/01/85	CB2	NO HIT			
DUWRIV2	DR-15	07/01/85	CB3	NO HIT			
DUWRIV2	DR-16	07/01/85	CB4	HIT			
DUWRIV2	DR-17	07/01/85	CB5	NO HIT			
DUWRIV2	DR-18	07/01/85	CC1	NO HIT			
DUWRIV2	DR-19	07/01/85	CC2	NO HIT			
DUWRIV2	DR-20	07/01/85	CC3	NO HIT			
DUWRIV2	DR-21	07/01/85	CC4	NO HIT			
DUWRIV2	DR-22	07/01/85	CC5	NO HIT			
DUWRIV2	DR-23	07/01/85	CD1	NO HIT			
DUWRIV2	DR-24	07/01/85	CD2	NO HIT			

DUWRIV2	DR-25	07/01/85	CE1	HIT
DUWRIV2	DR-26	07/01/85	CE2	HIT
DUWRIV2	DR-27	07/01/85	CE3	HIT
DUWRIV2	DR-28	07/01/85	CF1	NO HIT
DUWRIV2	DR-29	07/01/85	CF2	HIT
DUWRIV2	DR-30	07/01/85	CF3	NO HIT
DUWRIV2	DR-31	07/01/85	CF4	NO HIT
DUWRIV2	DR-32	07/01/85	CF5	NO HIT
DUWRIV2	DR-33	07/01/85	CG1	NO HIT
DUWRIV2	DR-34	07/01/85	CG2	NO HIT
DUWRIV2	DR-35	07/01/85	CG3	NO HITa
DUWRIV2	DR-36	07/01/85	CG4	NO HIT
DUWRIV2	DR-37	07/01/85	CG5	NO HIT
DUWRIV2	DR-38	07/01/85	CH1	NO HIT
DUWRIV2	DR-39	07/01/85	CW/A1	NO HIT
DUWRIV2	SQ-21	07/01/85	SEQUIM	NO HIT

*Amphipod toxicity was the only biological effect tested in the DUWRIV1 and DUWRIV2 surveys.

EBCHEM	AB-01	09/26/85	AB-01	HIT	b
EBCHEM	AB-02	09/26/85	AB-02	NO HIT	NO HIT
EBCHEM	AB-03	09/26/85	AB-03	NO HIT	HIT
EBCHEM	AB-04	09/26/85	AB-04	NO HIT	NO HIT
EBCHEM	DR-01	09/30/85	DR-01	NO HIT	
EBCHEM	DR-02	09/30/85	DR-02	HIT	
EBCHEM	DR-03	09/30/85	DR-03	NO HIT	
EBCHEM	DR-04	09/30/85	DR-04	NO HIT	
EBCHEM	DR-05	09/30/85	DR-05	c	
EBCHEM	DR-06	10/09/85	DR-06	NO HIT	
EBCHEM	DR-07	09/30/85	DR-07	NO HIT	
EBCHEM	DR-08	09/30/85	DR-08	c	
EBCHEM	DR-09	09/30/85	DR-09	NO HIT	
EBCHEM	DR-10	09/30/85	DR-10	b	
EBCHEM	DR-11	09/30/85	DR-11	NO HIT	
EBCHEM	DR-12	09/30/85	DR-12	b	
EBCHEM	DR-13	09/30/85	DR-13	HIT	
EBCHEM	DR-14	09/30/85	DR-14	HIT	
EBCHEM	DR-15	09/30/85	DR-15	HIT	
EBCHEM	DR-16	09/30/85	DR-16	HIT	
EBCHEM	DR-17	09/30/85	DR-17	NO HIT	
EBCHEM	DR-25	10/10/85	DR-25	HIT	
EBCHEM	EW-01	10/09/85	EW-01	NO HIT	
EBCHEM	EW-02	10/04/85	EW-02	HIT	HIT
EBCHEM	EW-03	10/04/85	EW-03	c	NO HIT
EBCHEM	EW-04	10/14/85	EW-04	HIT	HIT
EBCHEM	EW-05	10/14/85	EW-05	HIT	HIT
EBCHEM	EW-05	10/14/85	EW-05	HIT	HIT
EBCHEM	EW-06	10/04/85	EW-06	HIT	HIT
EBCHEM	EW-07	10/14/85	EW-07	HIT	HIT
EBCHEM	EW-08	10/14/85	EW-08	HIT	HIT
EBCHEM	EW-09	10/14/85	EW-09	HIT	HIT
EBCHEM	EW-10	10/14/85	EW-10	HIT	NO HIT
EBCHEM	EW-11	10/14/85	EW-11	HIT	HIT
EBCHEM	EW-12	10/15/85	EW-12	NO HIT	NO HIT
EBCHEM	EW-13	10/15/85	EW-13	NO HIT	NO HIT
EBCHEM	EW-14	10/15/85	EW-14	NO HIT	NO HIT

EBCHEM	EW-15	10/15/85	EW-15	NO HIT	NO HIT
EBCHEM	EW-16	10/15/85	EW-16	HIT	NO HIT
EBCHEM	KG-01	09/25/85	KG-01	NO HIT	HIT
EBCHEM	KG-02	10/09/85	KG-02	HIT	
EBCHEM	KG-03	09/25/85	KG-03	HIT	NO HIT
EBCHEM	KG-04	10/09/85	KG-04	NO HIT	
EBCHEM	KG-05	09/30/85	KG-05	HIT	HIT
EBCHEM	KG-06	09/30/85	KG-06	NO HIT	HIT
EBCHEM	KG-07	09/30/85	KG-07	NO HIT	HIT
EBCHEM	KG-08	10/01/85	KG-08	NO HIT	HIT
EBCHEM	KG-09	10/01/85	KG-09	HIT	NO HIT
EBCHEM	KG-10	10/08/85	KG-10	HIT	
EBCHEM	KG-11	10/01/85	KG-11	HIT	HIT
EBCHEM	MG-01	09/26/85	MG-01	NO HIT	HIT
EBCHEM	MG-02	09/26/85	MG-02	NO HIT	HIT
EBCHEM	MG-03	09/26/85	MG-03	NO HIT	HIT
EBCHEM	MG-04	09/26/85	MG-04	NO HIT	HIT
EBCHEM	NH-01	10/15/85	NH-01	NO HIT	HIT
EBCHEM	NH-02	10/15/85	NH-02	HIT	HIT
EBCHEM	NH-03	10/16/85	NH-03	HIT	HIT
EBCHEM	NH-04	10/15/85	NH-04	HIT	HIT
EBCHEM	NH-05	10/15/85	NH-05	HIT	HIT
EBCHEM	NH-06	10/16/85	NH-06	HIT	HIT
EBCHEM	NH-07	10/09/85	NH-07	NO HIT	
EBCHEM	NH-08	10/16/85	NH-08	HIT	HIT
EBCHEM	NH-09	10/16/85	NH-09	HIT	NO HIT
EBCHEM	NH-10	10/08/85	NH-10	NO HIT	
EBCHEM	NH-11	10/15/85	NH-11	HIT	HIT
EBCHEM	NS-01	10/08/85	NS-01	HIT	
EBCHEM	NS-02	09/27/85	NS-02	NO HIT	NO HIT
EBCHEM	NS-03	10/04/85	NS-03	NO HIT	NO HIT
EBCHEM	NS-04	10/08/85	NS-04	c	
EBCHEM	NS-05	10/04/85	NS-05	NO HIT	NO HIT
EBCHEM	NS-06	09/27/85	NS-06	b	HIT
EBCHEM	NS-07	10/04/85	NS-07	HIT	NO HIT
EBCHEM	NS-08	09/26/85	NS-08	HIT	HIT
EBCHEM	PS-01	10/12/85	PS-01	NO HIT	NO HIT
EBCHEM	PS-01	10/12/85	PS-01	NO HIT	NO HIT
EBCHEM	PS-02	10/12/85	PS-02	NO HIT	NO HIT
EBCHEM	PS-03	10/12/85	PS-03	NO HIT	NO HIT
EBCHEM	PS-04	10/12/85	PS-04	NO HIT	NO HIT
EBCHEM	SS-01	10/16/85	SS-01	NO HIT	NO HIT
EBCHEM	SS-03	10/04/85	SS-03	HIT	HIT
EBCHEM	SS-04	10/04/85	SS-04	b	HIT
EBCHEM	SS-05	10/03/85	SS-05	NO HIT	NO HIT
EBCHEM	SS-05	10/03/85	SS-05	NO HIT	NO HIT
EBCHEM	SS-06	10/03/85	SS-06	HIT	HIT
EBCHEM	SS-07	10/03/85	SS-07	HIT	NO HIT
EBCHEM	SS-08	09/27/85	SS-08	HIT	HIT
EBCHEM	SS-09	09/27/85	SS-09	HIT	HIT
EBCHEM	SS-10	09/27/85	SS-10	b	b
EBCHEM	SS-11	09/27/85	SS-11	NO HIT	NO HIT
EBCHEM	SS-12	09/27/85	SS-12	NO HIT	NO HIT
EBCHEM	WW-01	10/01/85	WW-01	NO HIT	HIT
EBCHEM	WW-02	10/09/85	WW-02	HIT	
EBCHEM	WW-03	10/01/85	WW-03	NO HIT	HIT

EBCHEM	WW-04	10/01/85	WW-04	NO HIT	HIT
EBCHEM	WW-05	10/01/85	WW-05	NO HIT	HIT
EBCHEM	WW-06	10/01/85	WW-06	NO HIT	HIT
EBCHEM	WW-06	10/01/85	WW-06	NO HIT	HIT
EBCHEM	WW-08	10/01/85	WW-08	HIT	HIT
EBCHEM	WW-09	10/02/85	WW-09	HIT	HIT
EBCHEM	WW-10	10/02/85	WW-10	NO HIT	HIT
EBCHEM	WW-11	10/02/85	WW-11	HIT	HIT
EBCHEM	WW-12	10/02/85	WW-12	HIT	HIT
EBCHEM	WW-13	10/02/85	WW-13	NO HIT	NO HIT
EBCHEM	WW-14	10/02/85	WW-14	b	HIT
EBCHEM	WW-15	10/08/85	WW-15	NO HIT	
EBCHEM	WW-16	10/02/85	WW-16	NO HIT	HIT
EBCHEM	WW-17	10/03/85	WW-17	NO HIT	HIT
EBCHEM	WW-18	10/03/85	WW-18	NO HIT	HIT
EBCHEM	WW-19	10/03/85	WW-19	NO HIT	HIT
EBCHEM	WW-20	10/03/85	WW-20	NO HIT	HIT
EHCHEM	BH-01	06/01/85	B1	NO HIT	NO HIT
EHCHEM	BH-02	06/01/85	B1	NO HIT	NO HIT
EHCHEM	EH-01	06/01/85	B1	NO HIT	NO HIT
EHCHEM	EH-02	06/01/85	B1	NO HIT	HIT
EHCHEM	EH-03	06/01/85	V6	NO HIT	HIT
EHCHEM	EH-05	06/01/85	V1	NO HIT	HIT
EHCHEM	EH-06	06/01/85	V6	NO HIT	HIT
EHCHEM	EH-08	06/01/85	V1	HIT	HIT
EHCHEM	EH-15	06/01/85	V1	NO HIT	HIT
EHCHEM	EH-16	06/01/85	V1	NO HIT	HIT

*Amphipod toxicity and benthic infauna abundance were the only biological effects tested in the EBCHEM and EHCHEM surveys.

EIGHTBAY	BH-03	01/01/82	BH-03	NO HIT
EIGHTBAY	BH-04	01/01/82	BH-04	NO HIT
EIGHTBAY	BH-05	01/01/82	BH-05	HIT
EIGHTBAY	BH-07	01/01/82	BH-07	NO HIT
EIGHTBAY	BH-11	01/01/82	BH-11	NO HIT
EIGHTBAY	BH-12	01/01/82	BH-12	NO HIT
EIGHTBAY	BH-23	01/01/82	BH-23	HIT
EIGHTBAY	BH-24	01/01/82	BH-24	NO HIT
EIGHTBAY	CS-01	01/01/82	CS-01	HIT
EIGHTBAY	CS-11	01/01/82	CS-11	HIT
EIGHTBAY	CS-15	01/01/82	CS-15	HIT
EIGHTBAY	CS-17	01/01/82	CS-17	HIT
EIGHTBAY	DB-01	01/01/82	DB-01	NO HIT
EIGHTBAY	DB-05	01/01/82	DB-05	NO HIT
EIGHTBAY	DB-07	01/01/82	DB-07	HIT
EIGHTBAY	DB-15	01/01/82	DB-15	HIT
EIGHTBAY	EL-09	01/01/82	EL-09	NO HIT
EIGHTBAY	EL-10	01/01/82	EL-10	NO HIT
EIGHTBAY	EL-12	01/01/82	EL-12	NO HIT
EIGHTBAY	EL-17	01/01/82	EL-17	NO HIT
EIGHTBAY	EL-20	01/01/82	EL-20	NO HIT
EIGHTBAY	EL-22	01/01/82	EL-22	NO HIT
EIGHTBAY	EL-23	01/01/82	EL-23	NO HIT
EIGHTBAY	EL-24	01/01/82	EL-24	NO HIT

EIGHTBAY	EV-01	01/01/82	EV-01	HIT	
EIGHTBAY	EV-02	01/01/82	EV-02	HIT	
EIGHTBAY	EV-03	01/01/82	EV-03	HIT	
EIGHTBAY	EV-04	01/01/82	EV-04	HIT	
EIGHTBAY	EV-05	01/01/82	EV-05	HIT	
EIGHTBAY	EV-06	01/01/82	EV-06	NO HIT	
EIGHTBAY	EV-07	01/01/82	EV-07	NO HIT	
EIGHTBAY	EV-11	01/01/82	EV-11	HIT	
EIGHTBAY	SC-06	01/01/82	SC-06	HIT	
EIGHTBAY	SC-07	01/01/82	SC-07	NO HIT	
EIGHTBAY	SC-08	01/01/82	SC-08	HIT	
EIGHTBAY	SC-14	01/01/82	SC-14	HIT	
EIGHTBAY	SC-17	01/01/82	SC-17	HIT	
EIGHTBAY	SC-18	01/01/82	SC-18	HIT	
EIGHTBAY	SC-19	01/01/82	SC-19	NO HITa	
EIGHTBAY	SC-20	01/01/82	SC-20	HIT	
EIGHTBAY	SM-01	01/01/82	SM-01	HIT	
EIGHTBAY	SM-03	01/01/82	SM-03	HIT	
EIGHTBAY	SM-07	01/01/82	SM-07	HIT	
EIGHTBAY	SM-20	01/01/82	SM-20	HIT	
EIGHTBAY	SQ-14	01/01/82	SQ-14	NO HIT	
EIGHTBAY	SQ-17	01/01/82	SQ-17	NO HIT	
EIGHTBAY	SQ-18	01/01/82	SQ-18	NO HIT	
EIGHTBAY	SQ-20	01/01/82	SQ-20	NO HIT	

*Amphipod toxicity was the only biological effect tested in the EIGHTBAY survey.

EVCHEM	ES-01	10/06/86	ES-01G	NO HIT	
EVCHEM	ES-02	10/06/86	ES-02G	NO HIT	
EVCHEM	ES-03	10/06/86	ES-03G	NO HIT	
EVCHEM	EW-01	10/07/86	EW-01G	HIT	HIT
EVCHEM	EW-04	09/30/86	EW-04G	HIT	HIT
EVCHEM	EW-07	09/30/86	EW-07G	HIT	HIT
EVCHEM	EW-10	10/01/86	EW-10G	HIT	HIT
EVCHEM	EW-12	10/01/86	EW-12G	NO HIT	HIT
EVCHEM	EW-14	10/01/86	EW-14G	c	HIT
EVCHEM	NG-01	10/02/86	NG-01G	NO HIT	NO HIT
EVCHEM	NG-02	10/02/86	NG-02G	NO HIT	HIT
EVCHEM	NG-03	10/02/86	NG-03G	NO HIT	HIT
EVCHEM	NG-04	10/02/86	NG-04G	HIT	HIT
EVCHEM	NG-06	10/08/86	NG-06G	HIT	NO HIT
EVCHEM	NG-10	10/03/86	NG-10G	NO HIT	NO HIT
EVCHEM	NG-12	10/15/86	NG-12G	NO HIT	
EVCHEM	NG-13	10/15/86	NG-13G	NO HIT	
EVCHEM	NG-14	10/15/86	NG-14G	NO HIT	
EVCHEM	NG-15	10/15/86	NG-15G	NO HIT	
EVCHEM	OG-03	10/09/86	OG-03G	HIT	
EVCHEM	PS-02	10/13/86	PS-02G	NO HIT	NO HIT
EVCHEM	PS-03	10/13/86	PS-03G	NO HIT	NO HIT
EVCHEM	PS-04	10/10/86	PS-04G	NO HIT	NO HIT
EVCHEM	SD-01	10/09/86	SD-01G	NO HIT	HIT
EVCHEM	SD-02	10/07/86	SD-02G	c	NO HIT
EVCHEM	SR-01	10/06/86	SR-01G	NO HIT	
EVCHEM	SR-02	10/06/86	SR-02G	NO HIT	
EVCHEM	SR-04	10/06/86	SR-04G	NO HIT	
EVCHEM	SR-07	10/03/86	SR-07G	c	HIT

EVCHEM	SR-08	10/03/86	SR-08G	NO HIT	HIT
EVCHEM	SS-01	10/06/86	SS-01G	NO HIT	
EVCHEM	SS-03	10/06/86	SS-03G	NO HIT	

*Amphipod toxicity and benthic infauna abundance were the only biological effects tested in the EVCHEM survey.

EVERETT1	EV-20	01/01/85	EV-20	NO HIT
EVERETT1	EV-21	01/01/85	EV-21	NO HIT
EVERETT1	EV-22	01/01/85	EV-22	NO HIT
EVERETT1	EV-23	01/01/85	EV-23	NO HIT
EVERETT1	EV-24	01/01/85	EV-24	HIT
EVERETT1	EV-25	01/01/85	EV-25	NO HIT

*Amphipod toxicity was the only biological effect tested in the EVERETT1 survey.

TPPS3AB	EB-33	03/15/82	1779	HIT
TPPS3AB	EB-33	07/15/82	2080	HIT
TPPS3AB	EB-35	03/15/82	1775	HIT
TPPS3AB	EB-35	07/15/82	2079	HIT
TPPS3AB	EB-36	03/15/82	1776	HIT
TPPS3AB	EB-36	07/15/82	2072	HIT
TPPS3AB	EB-38	03/15/82	1778	HIT
TPPS3AB	EB-38	07/15/82	2074	HIT
TPPS3AB	WP-01	07/15/82	2088	NO HIT
TPPS3AB	WP-02	07/15/82	2089	NO HIT
TPPS3AB	WP-03	07/15/82	2090	HIT
TPPS3AB	WP-04	07/15/82	2091	NO HIT
TPPS3AB	WP-05	07/15/82	2092	NO HIT
TPPS3AB	WP-06	07/15/82	2084	NO HIT
TPPS3AB	WP-07	07/15/82	2093	NO HIT
TPPS3AB	WP-08	07/15/82	2083	HIT
TPPS3AB	WP-09	07/15/82	2082	NO HIT
TPPS3AB	WP-10	07/15/82	2076	NO HIT
TPPS3AB	WP-11	03/15/82	1789	b
TPPS3AB	WP-12	03/15/82	1786	NO HIT
TPPS3AB	WP-12	07/15/82	2069	NO HIT
TPPS3AB	WP-13	03/15/82	1784	NO HIT
TPPS3AB	WP-13	07/15/82	2070	NO HIT
TPPS3AB	WP-14	03/15/82	1785	NO HIT
TPPS3AB	WP-14	07/15/82	2085	b
TPPS3AB	WP-15	03/15/82	1817	NO HIT
TPPS3AB	WP-15	07/15/82	2094	NO HIT
TPPS3AB	WP-16	03/15/82	1816	HIT
TPPS3AB	WP-16	07/15/82	2086	NO HITa

*Benthic infauna abundance was the only biological effect tested in the TPPS3AB survey.

---

a Although the indicated biological effect was statistically significant ($P \leq 0.05$) it is not classified as a "hit" because mortality was <25% for AMPT or abundance depression was <50% for BENA.

b The indicated biological effect was excluded because of a chemical anomaly.

c The indicated biological effect was excluded as inconclusive because of inadequate statistical power.

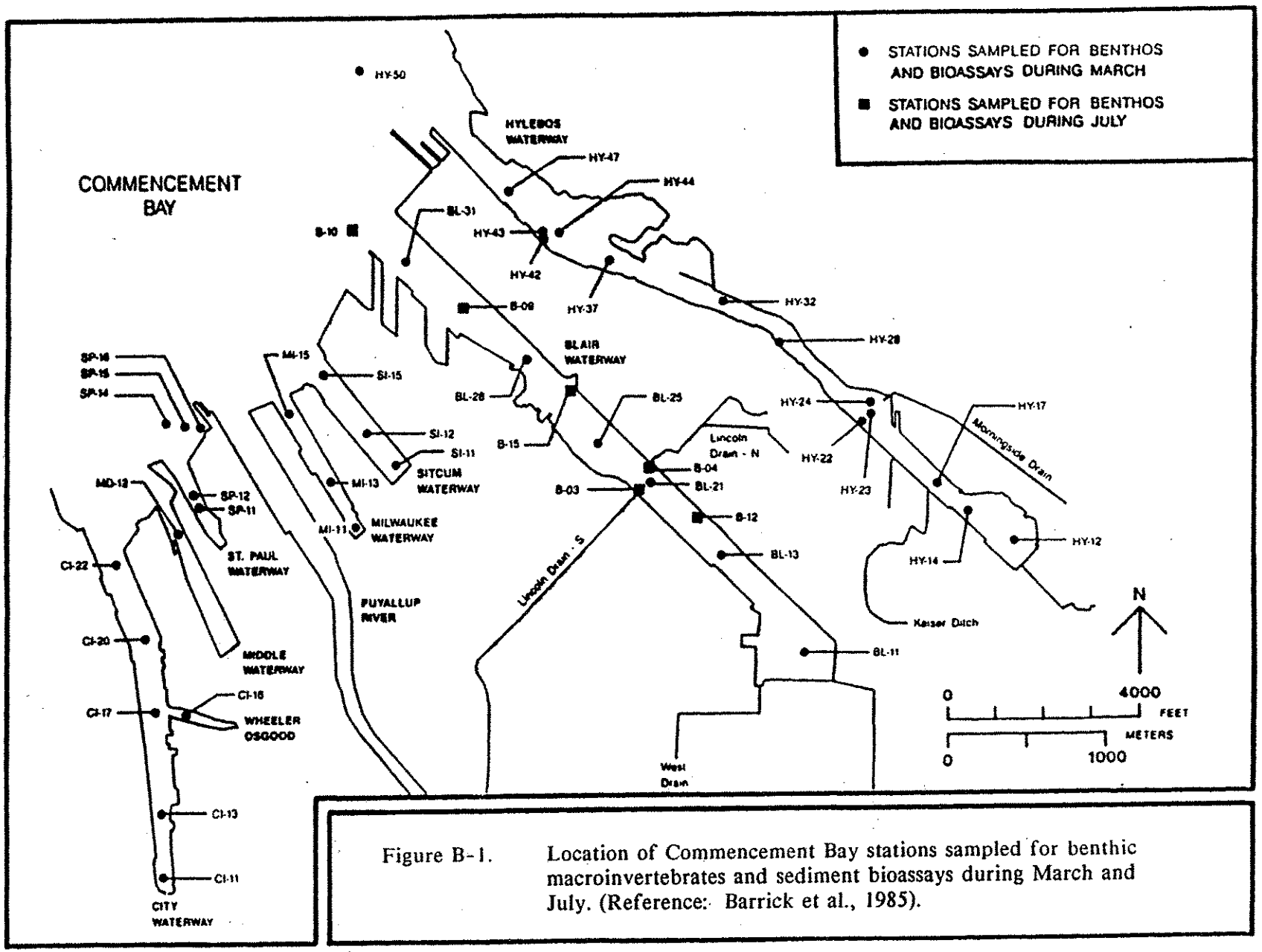


Figure B-1. Location of Commencement Bay stations sampled for benthic macroinvertebrates and sediment bioassays during March and July. (Reference: Barrick et al., 1985).

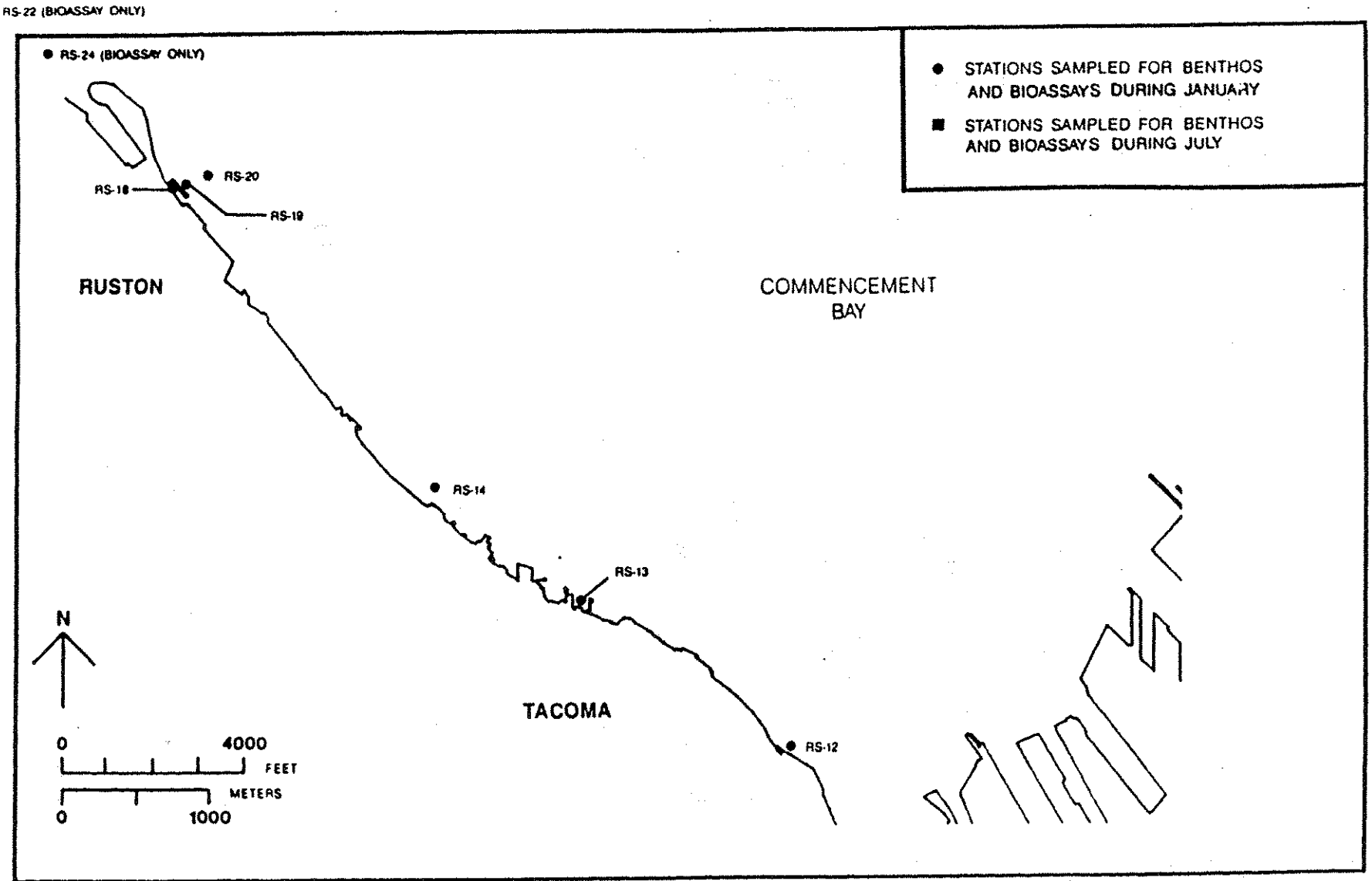


Figure B-1. (Continued).

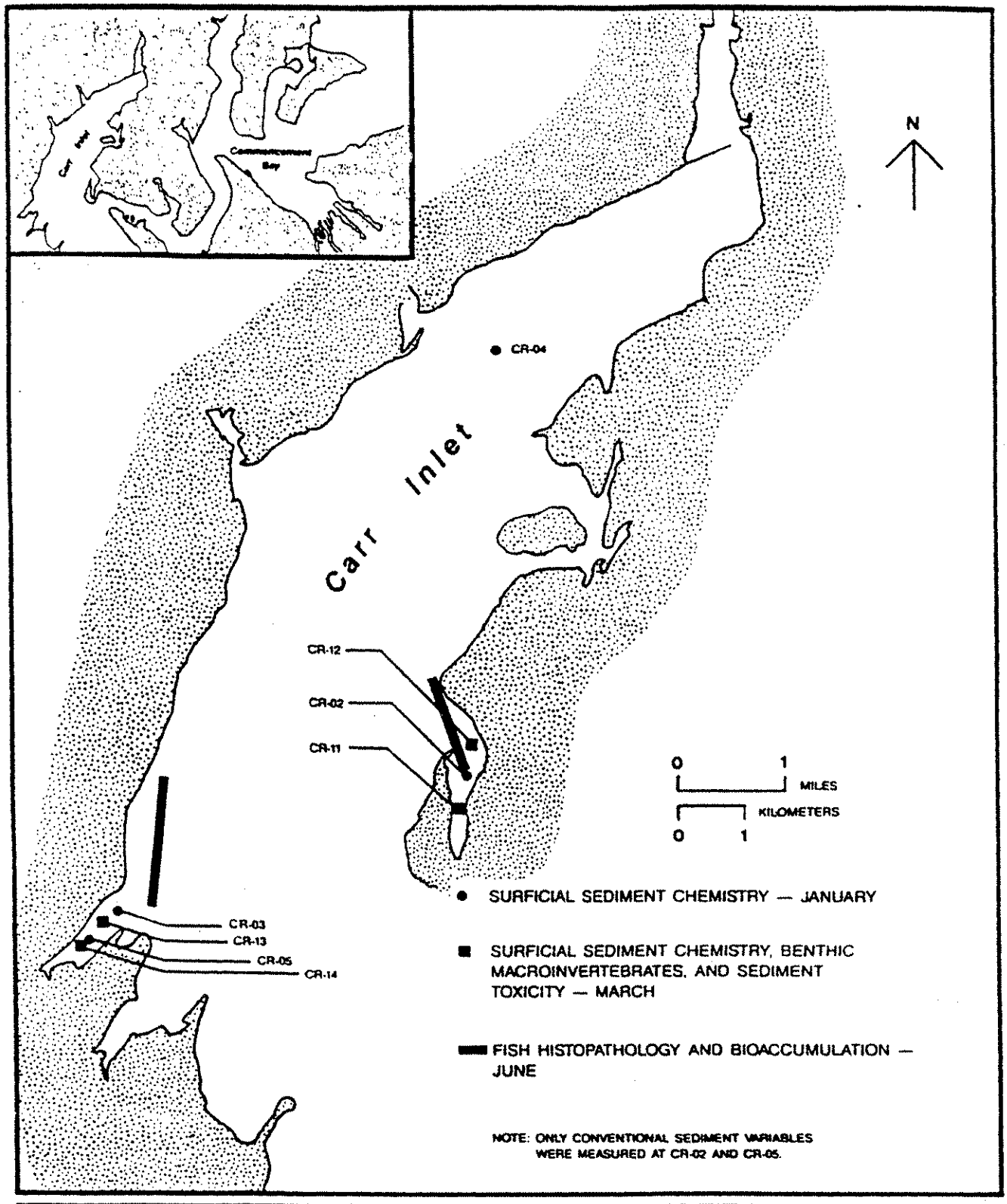


Figure B-2. Locations of reference stations sampled in Carr Inlet. (Reference: Barrick et al. 1985).

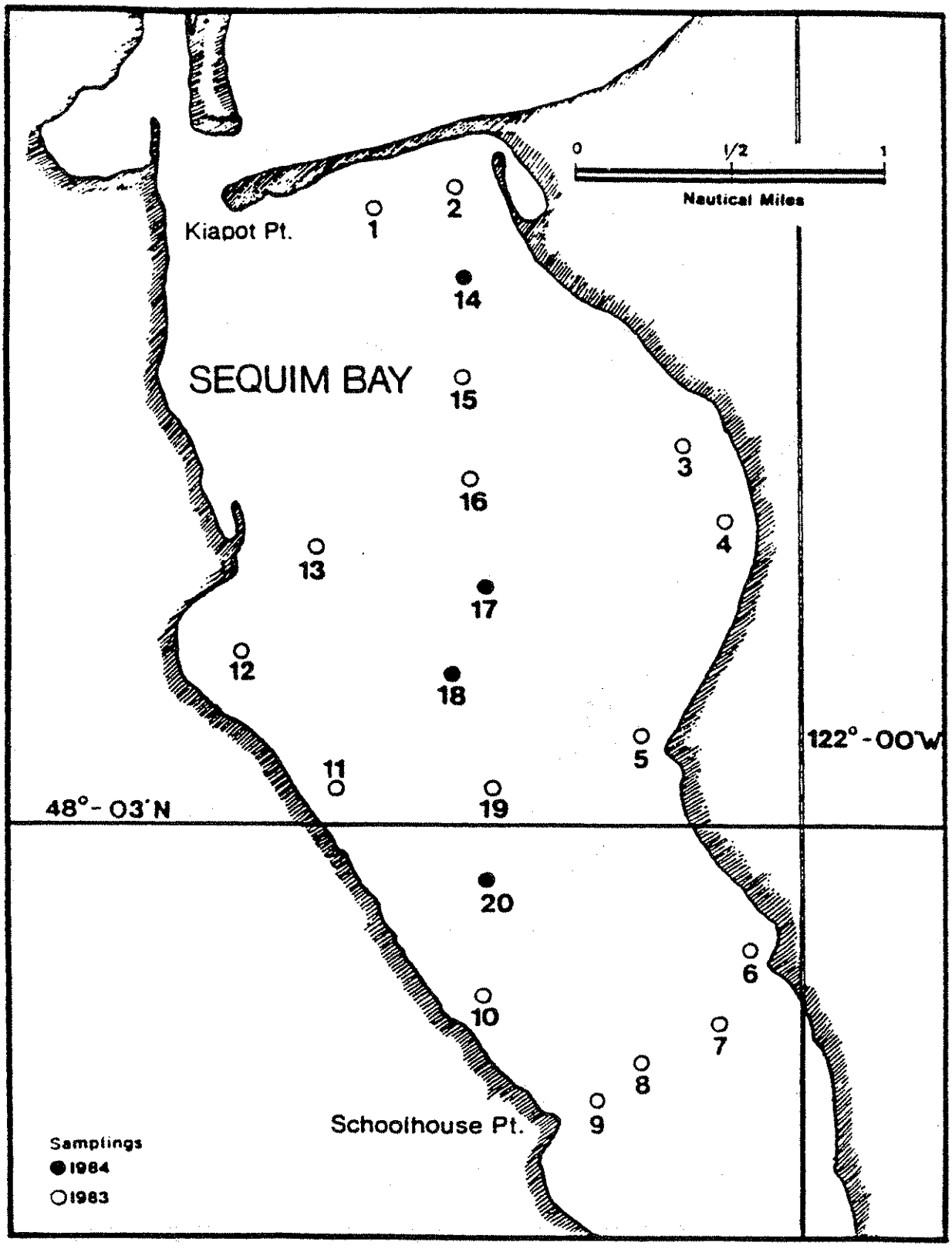


Figure B-3. Sequim Bay sampling stations. (Reference: Battelle 1985).

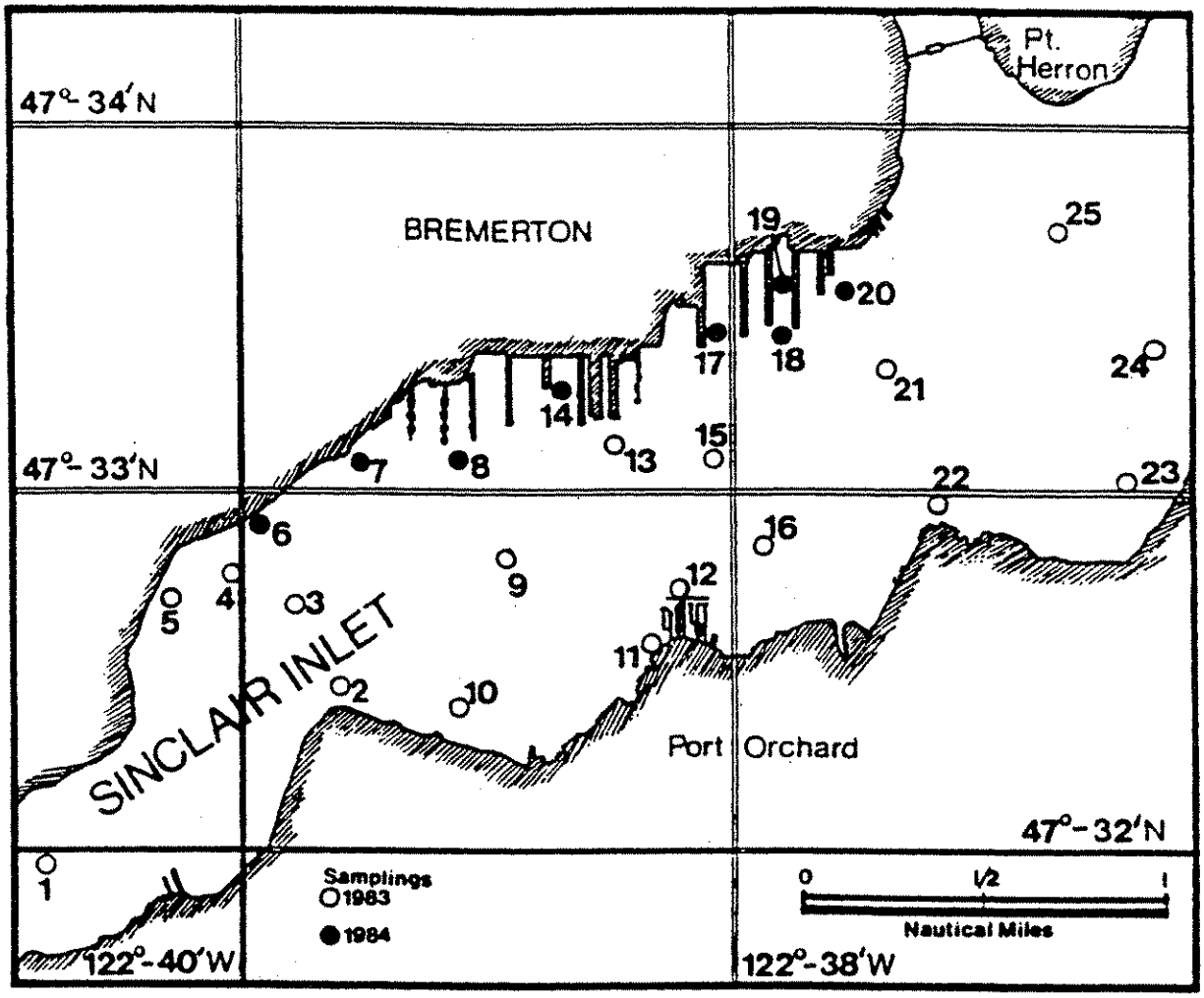


Figure B-4. Sinclair Inlet sampling stations. (Reference: Battelle 1985).

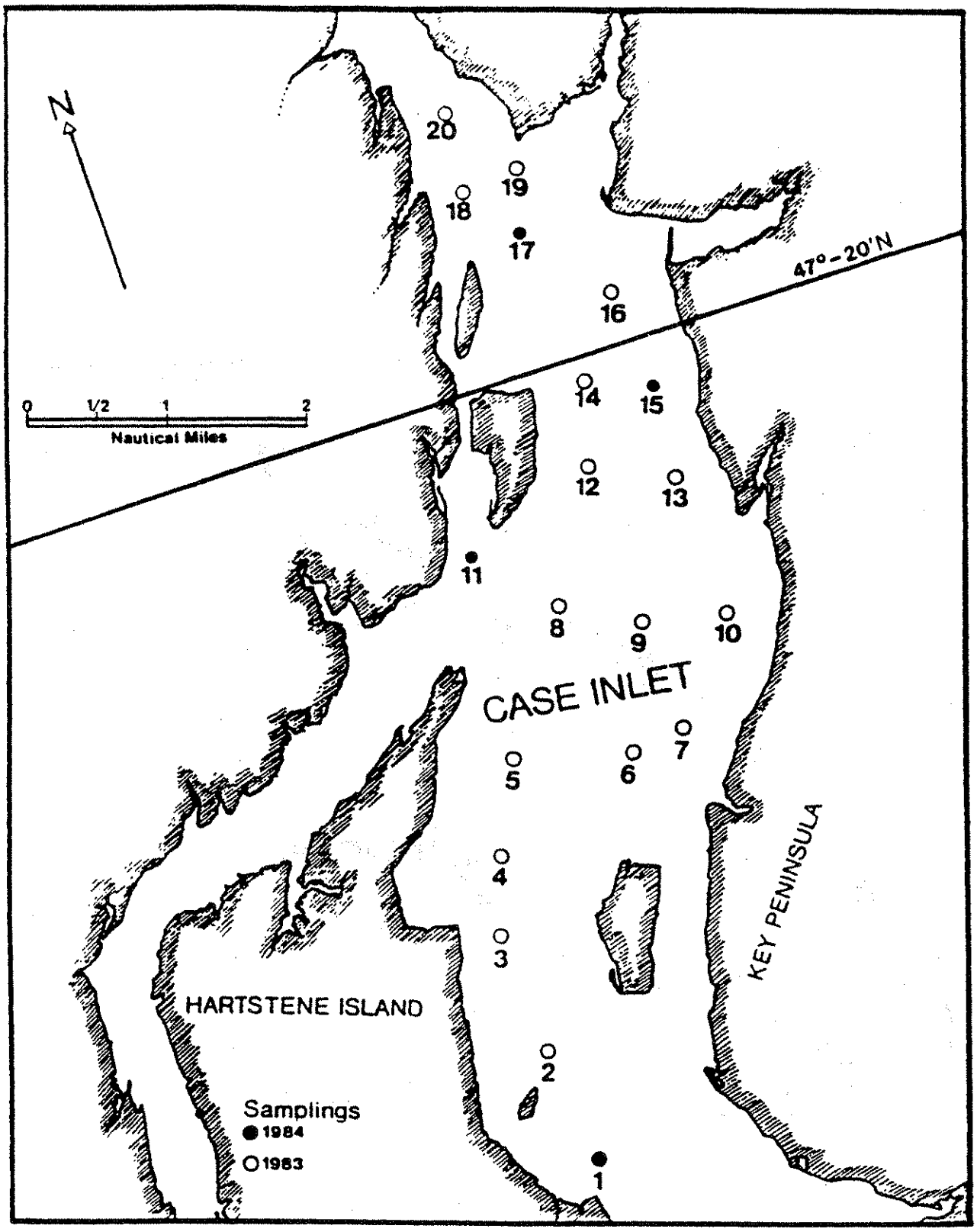


Figure B-5. Case Inlet sampling stations. (Reference: Battelle 1985).

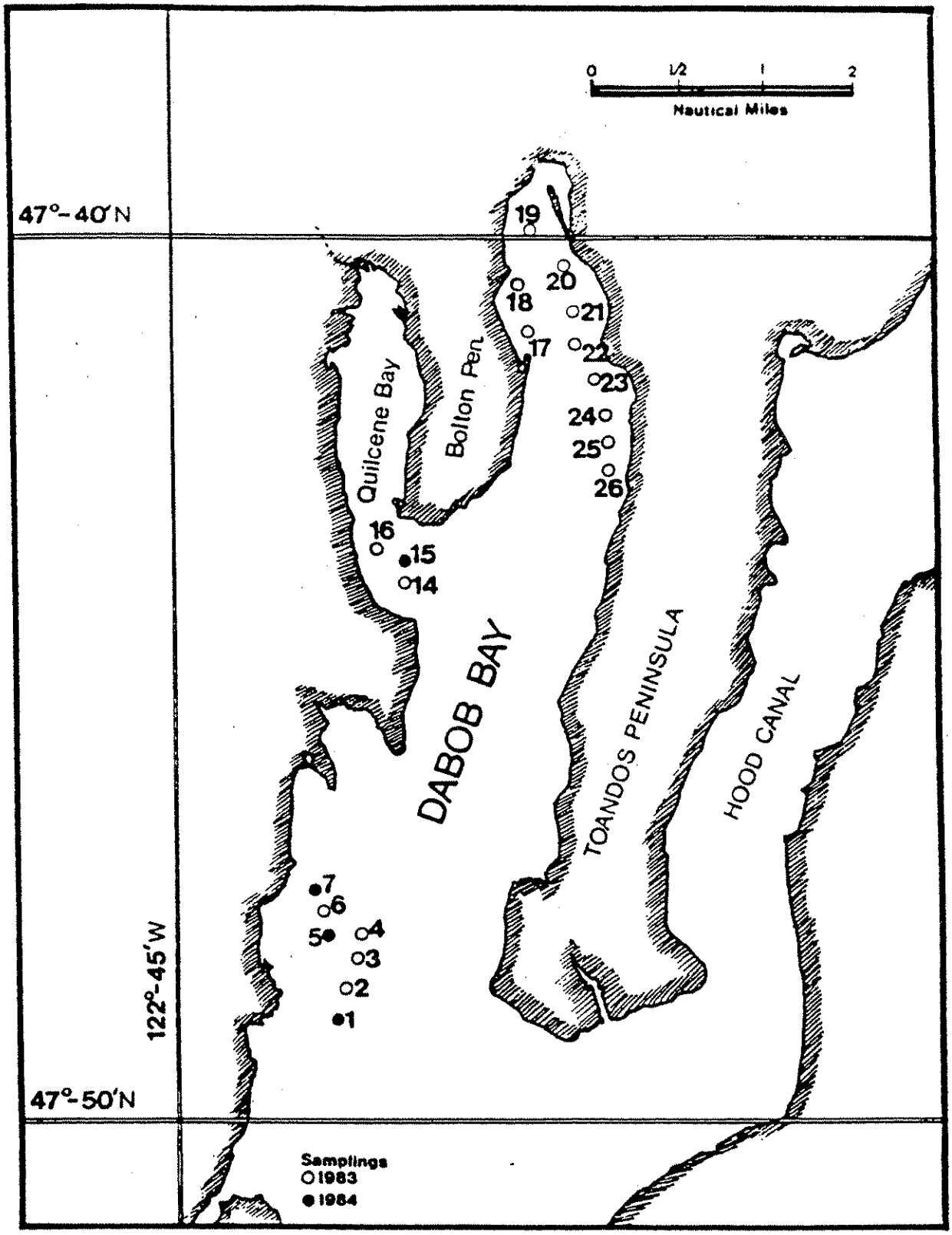


Figure B-6. Dabob Bay sampling stations. (Reference: Battelle 1985).

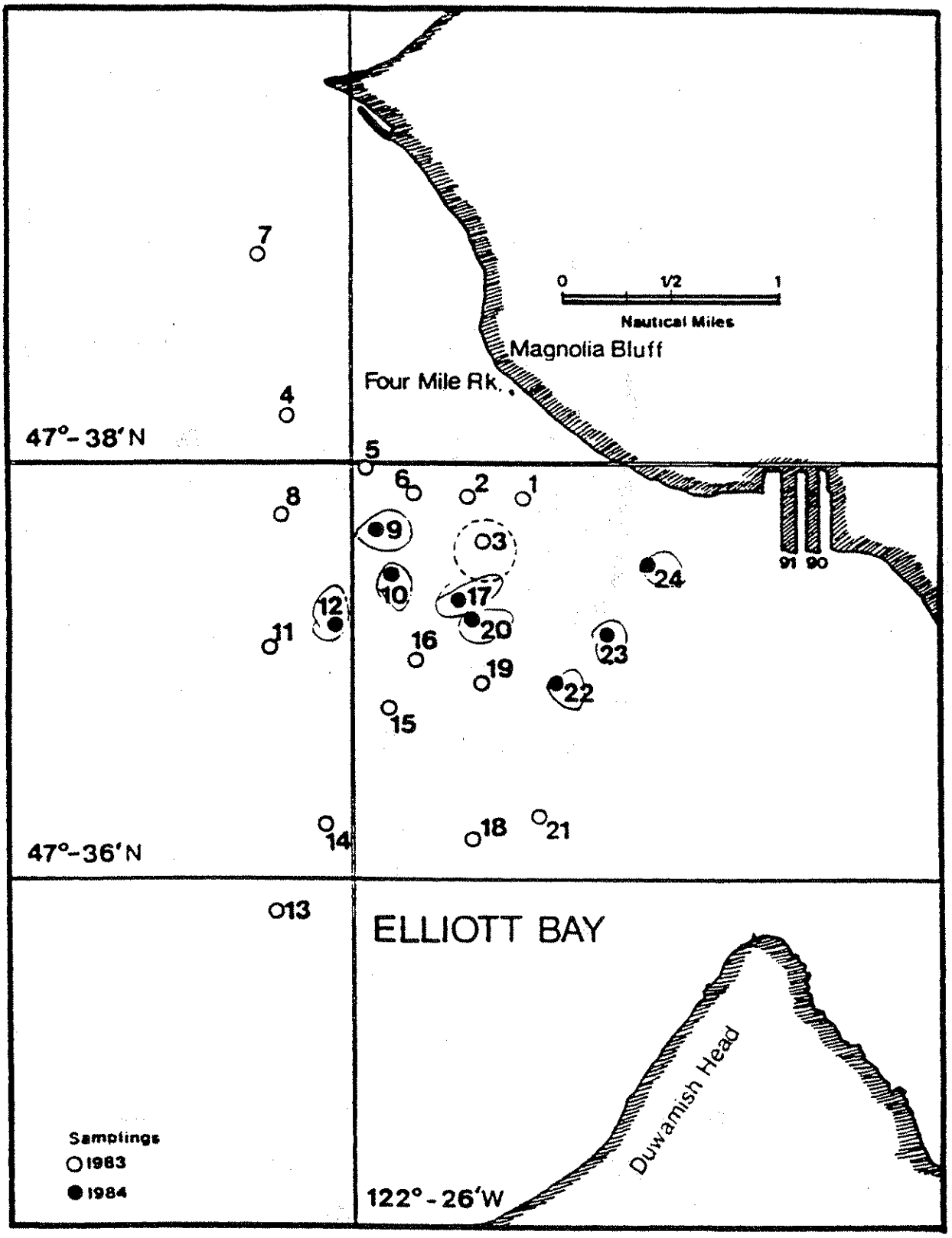


Figure B-7. Elliott Bay - Fourmile Rock sampling stations. (Reference: Battelle 1985).

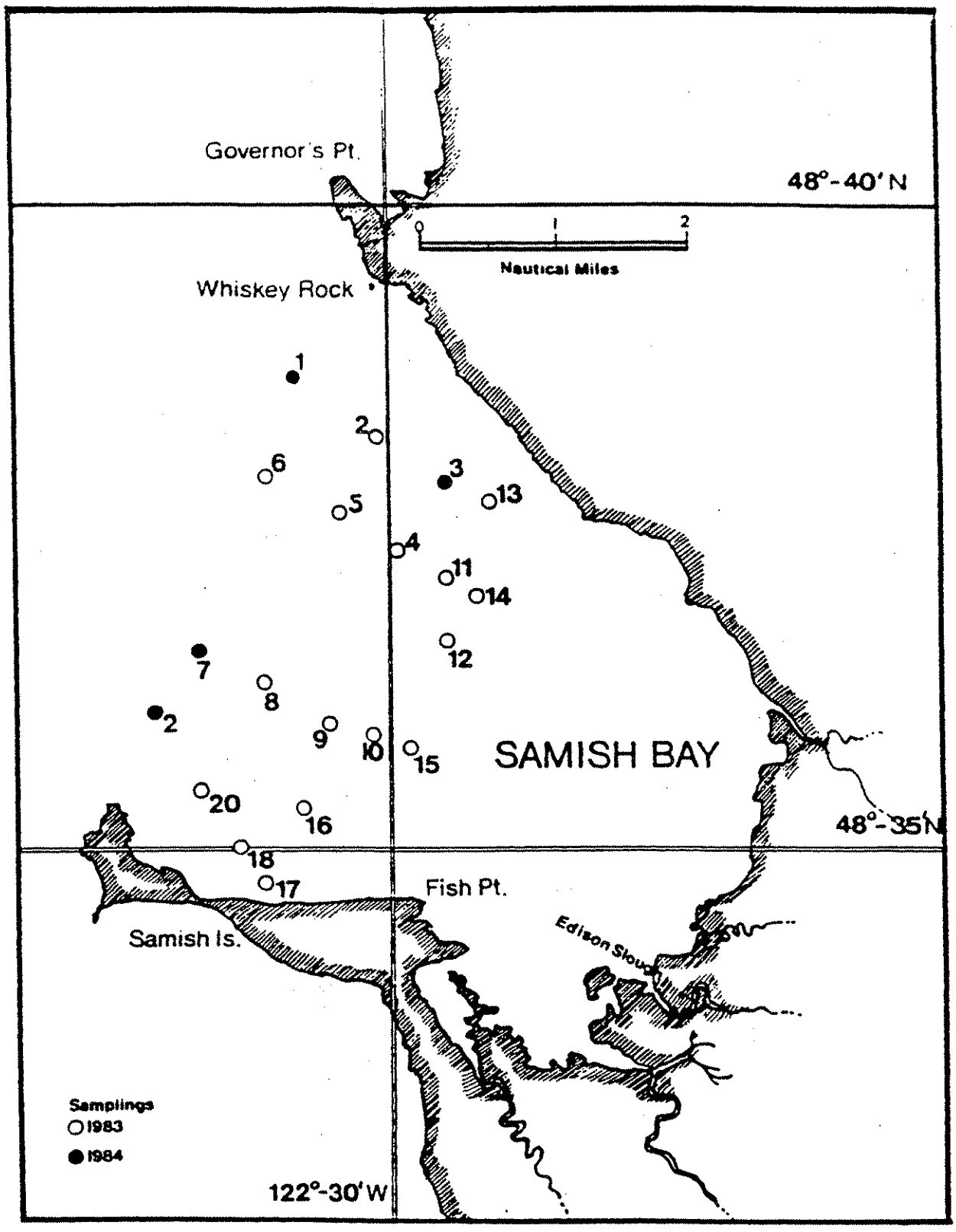


Figure B-8. Samish Bay sampling stations. (Reference: Battelle 1985).

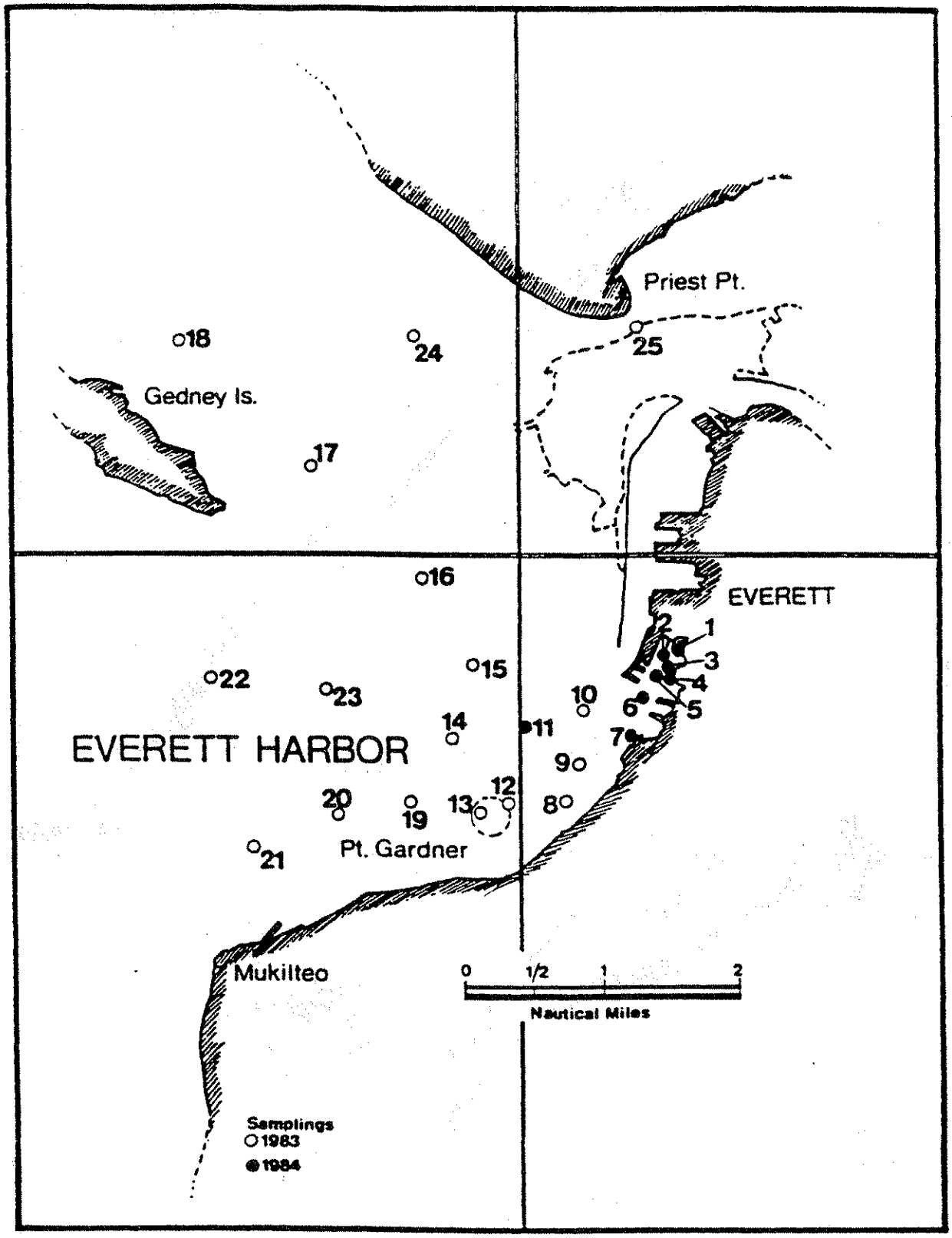


Figure B-9. Everett Harbor - Port Gardner sampling stations. (Reference: Battelle 1985).

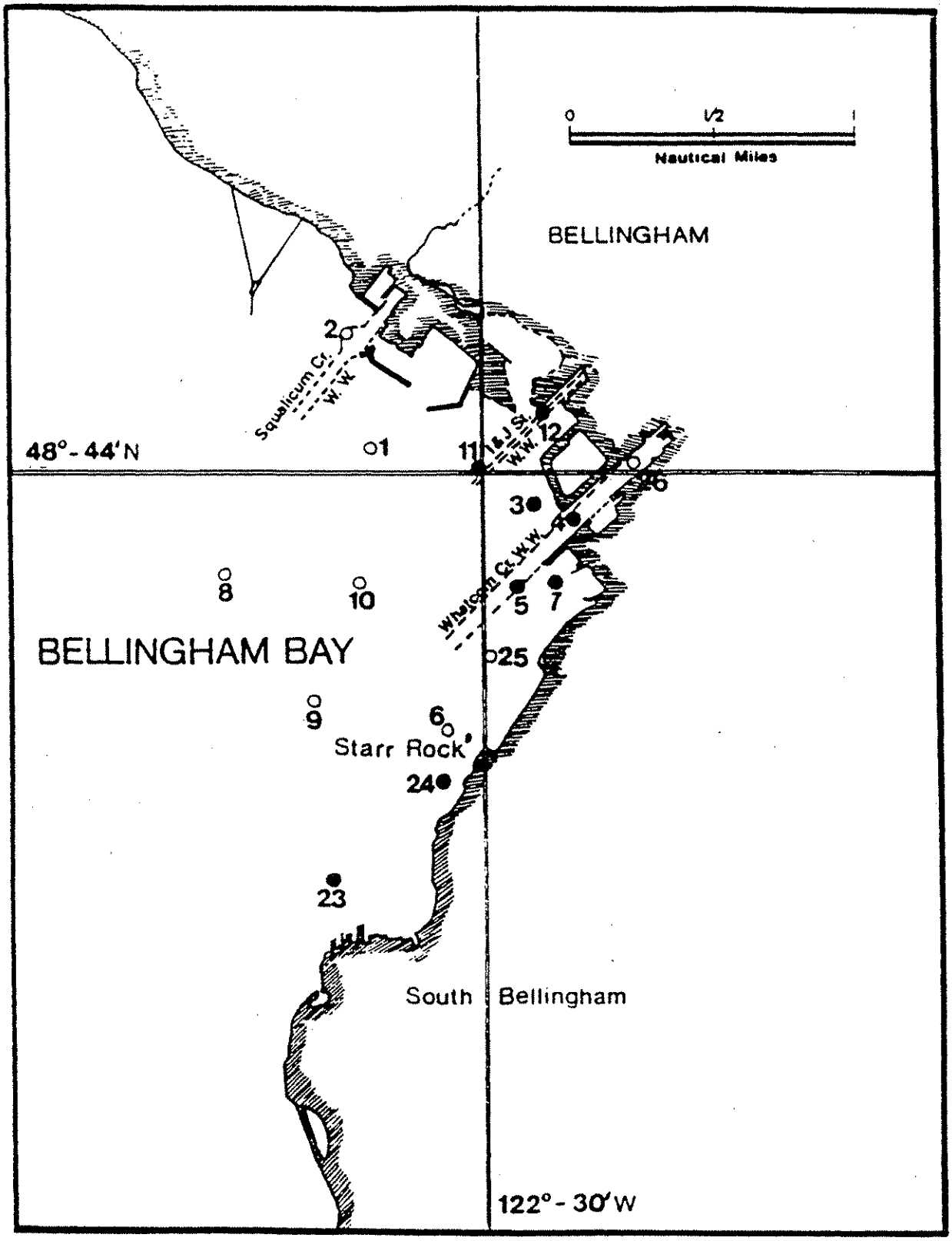


Figure B-10. Bellingham Bay sampling stations (Inner Harbor). (Reference: Battelle 1985).

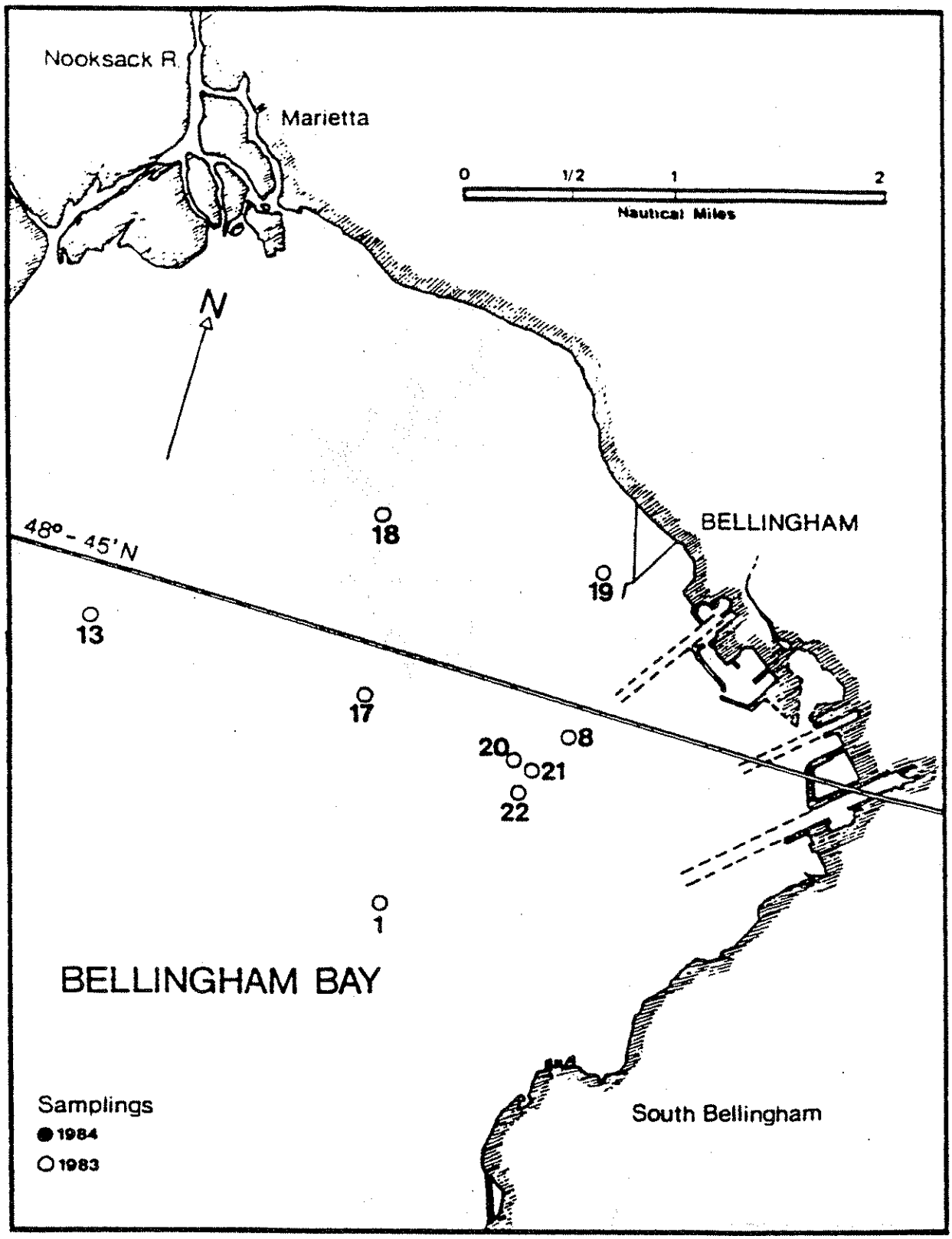


Figure B-11. Bellingham Bay sampling stations (Outer Harbor). (Reference: Battelle 1985).

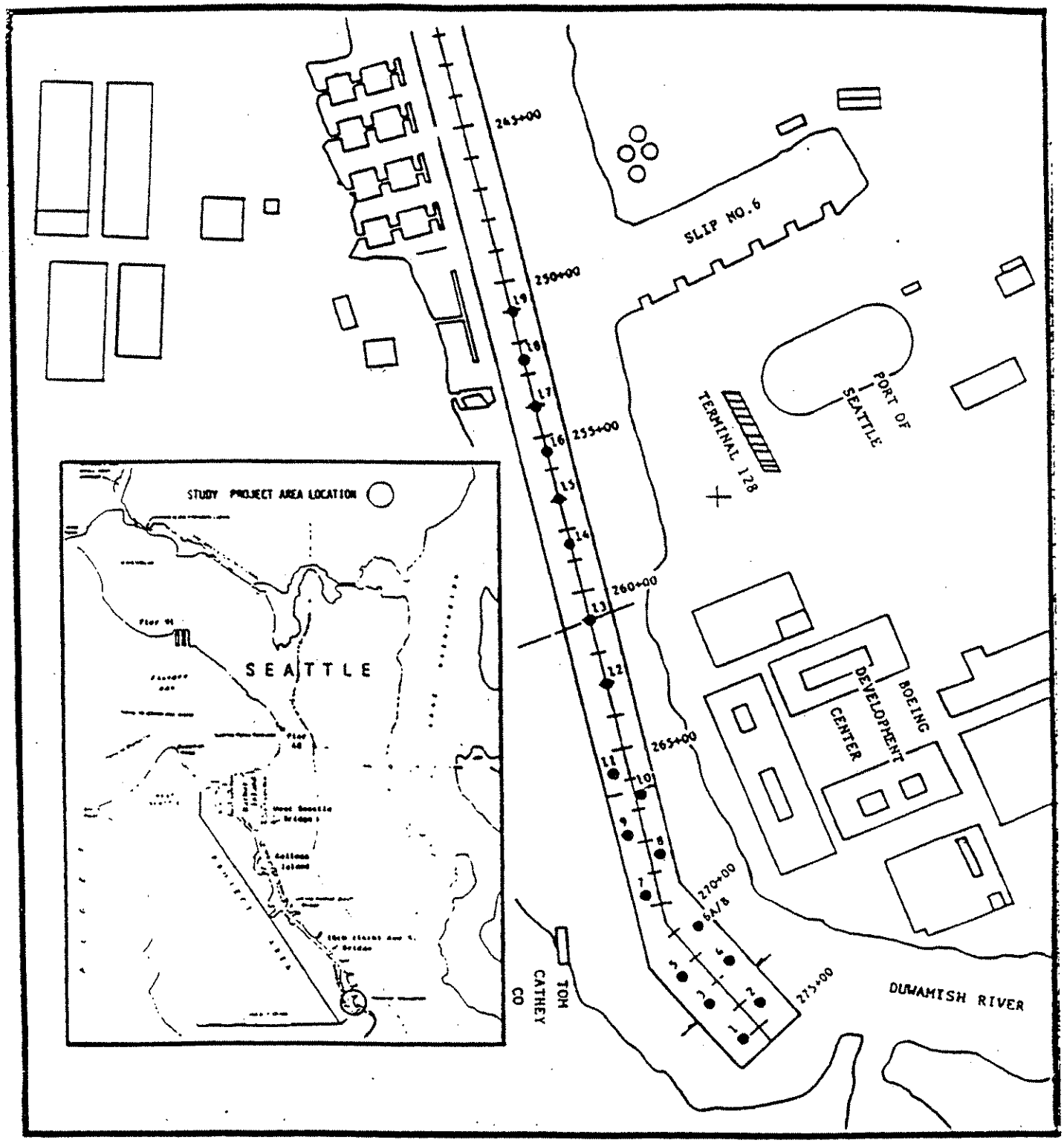


Figure B-12. Sediment sampling station locations for dredged material characterization in Duwamish River. (Reference: Chan et al. 1985).

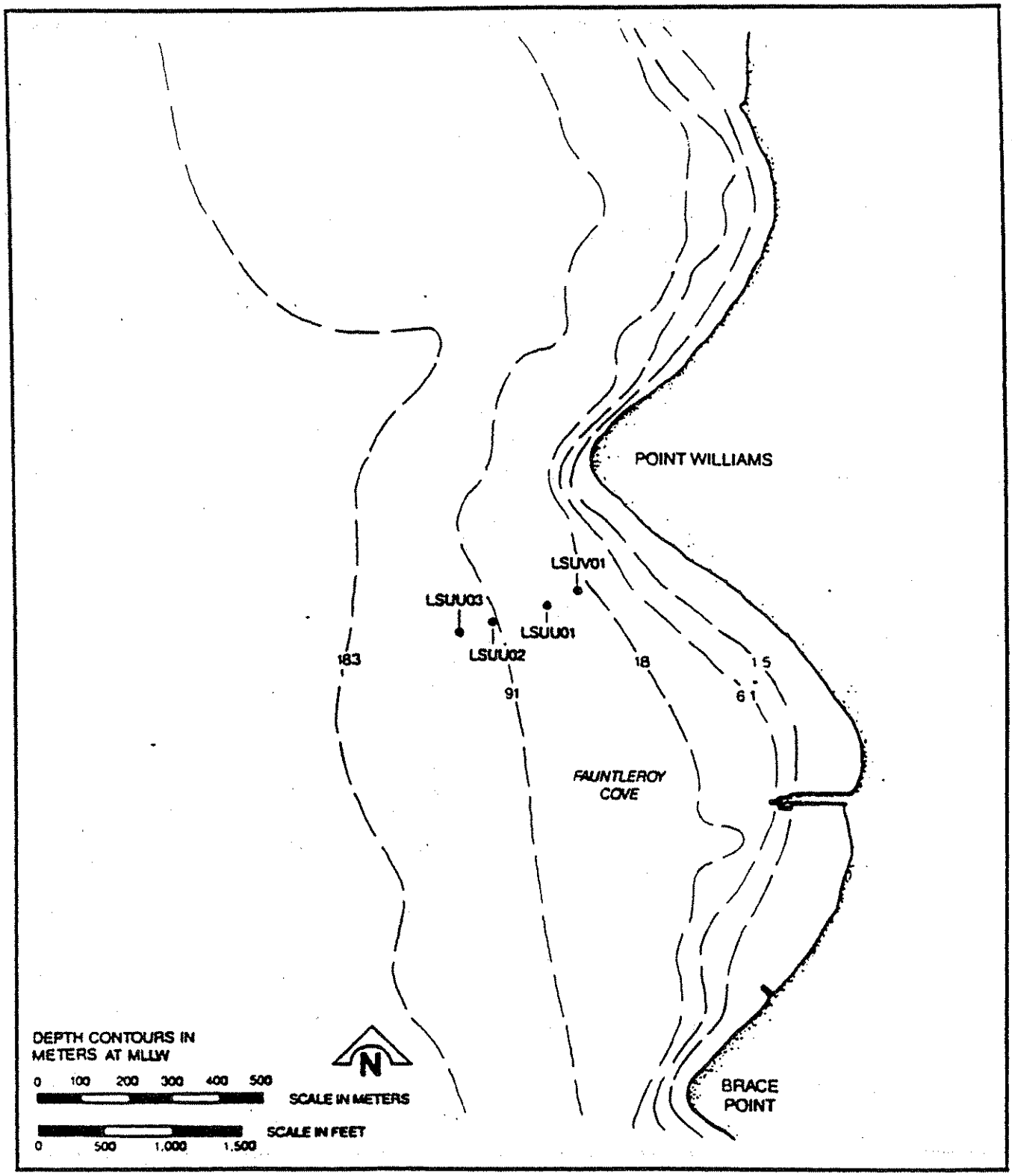


Figure B-13. Point Williams benthos reference sampling station locations. (Reference: Osborn et al. 1985).

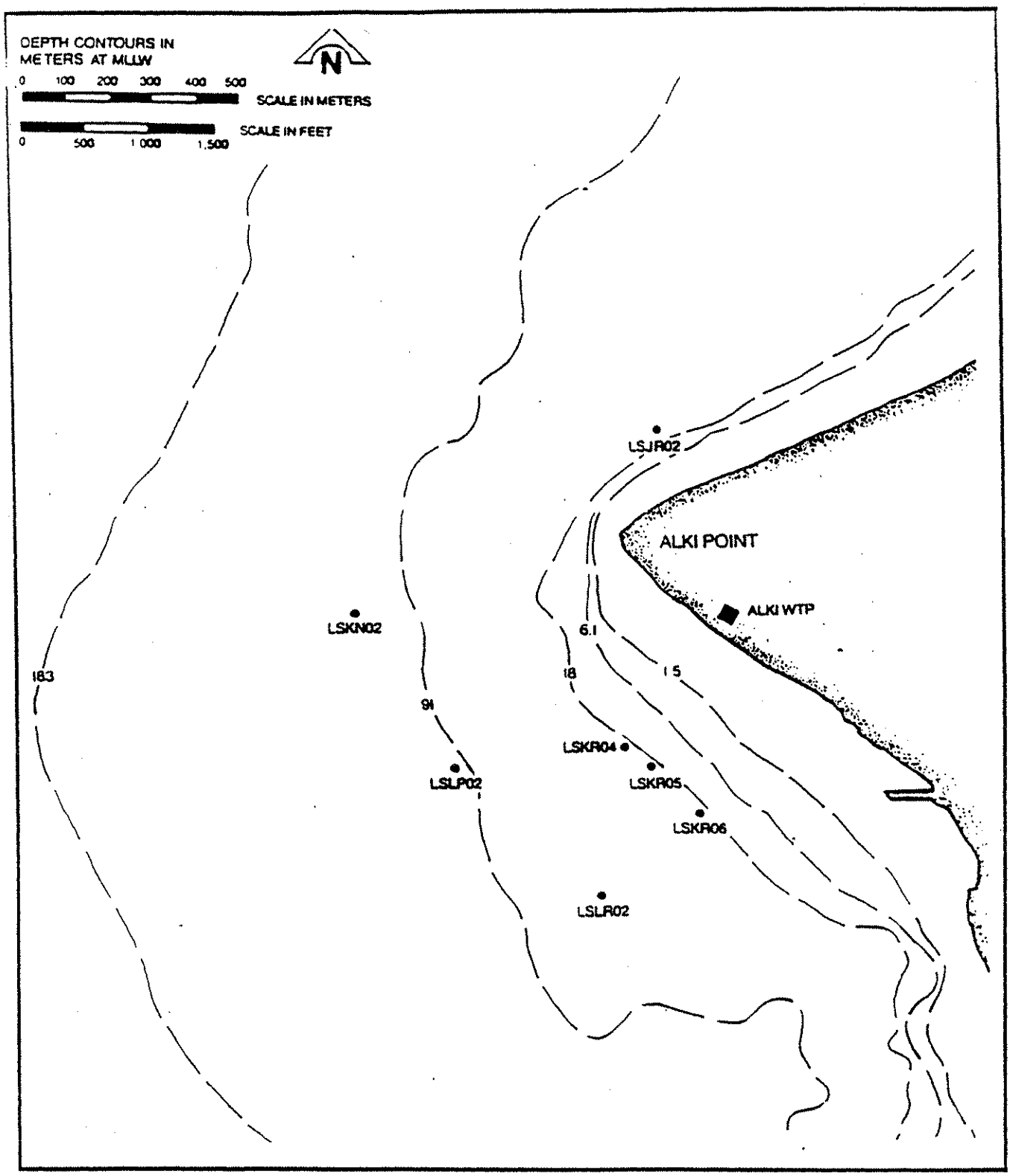


Figure B-14. Alki Point benthos sampling station locations. (Reference: Osborn et al. 1985).

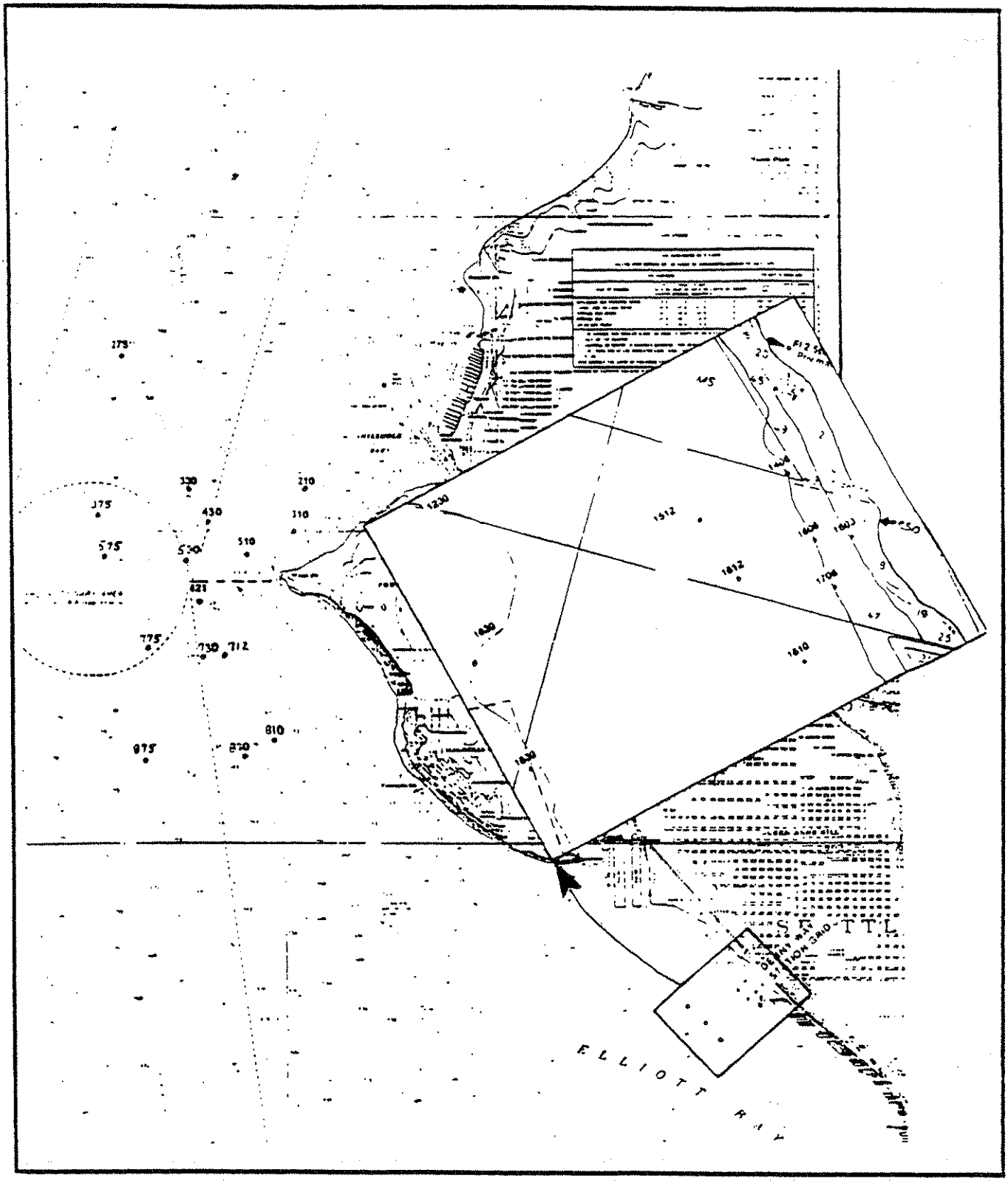


Figure B-15. Map showing the 26 stations in the central basin of Puget Sound and Elliott Bay sampled during Phase III of the TPPS program. (Reference: Romberg et al. 1984).

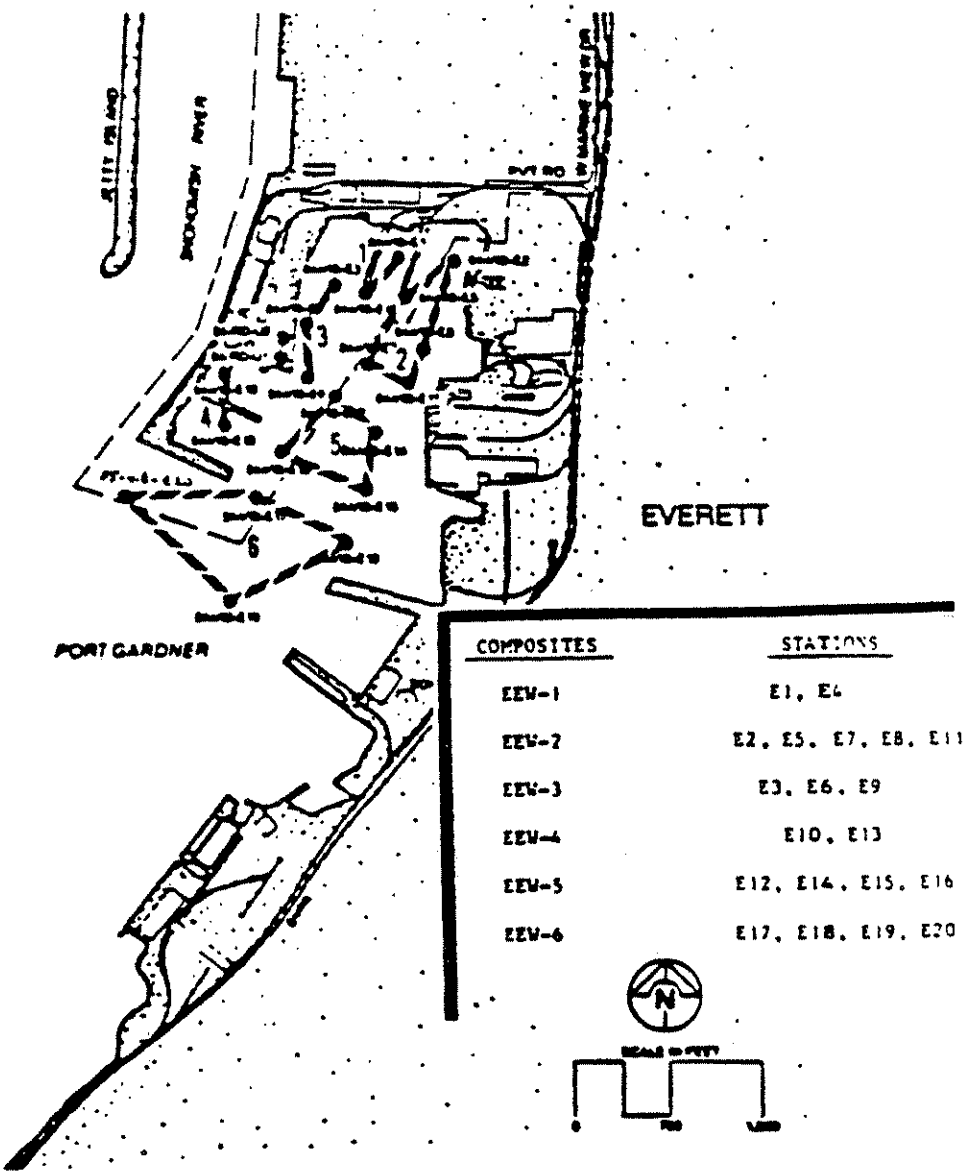


Figure B-16. Navy sediment sampling locations in the East Waterway of Everett. (Reference: U.S. Navy 1985).

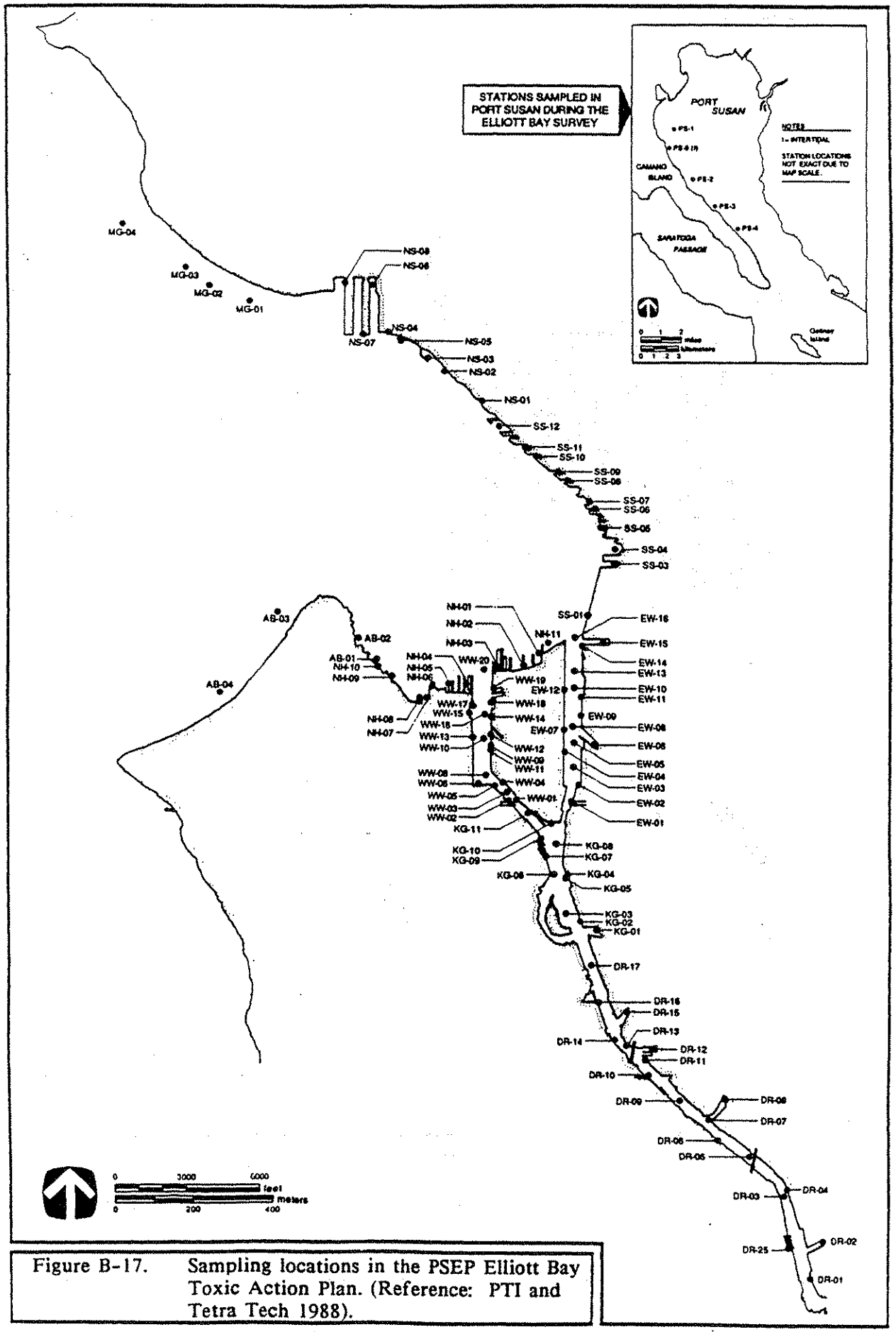


Figure B-17. Sampling locations in the PSEP Elliott Bay Toxic Action Plan. (Reference: PTI and Tetra Tech 1988).

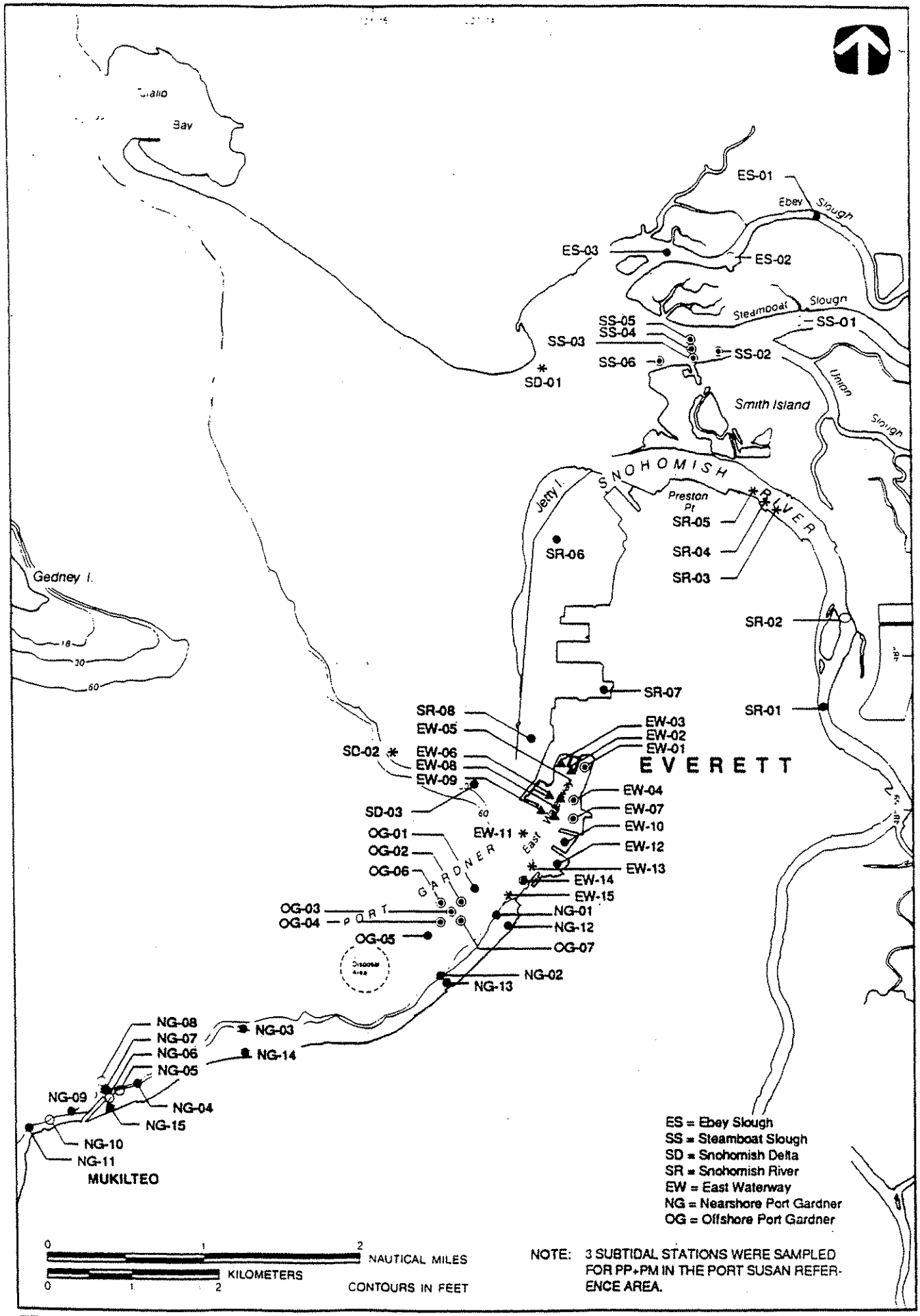


Figure B-18. Approximate sampling locations for PSEP chemical studies in Everett Harbor. (Reference: PTI and Tetra Tech 1988).

PP (INCLUDES VOLATILES)
PP - V
PP + PM
PP - V + PM
V + PM
PP = Priority Pollutants
V = Volatile Compounds
PM = Pulpmill Compounds

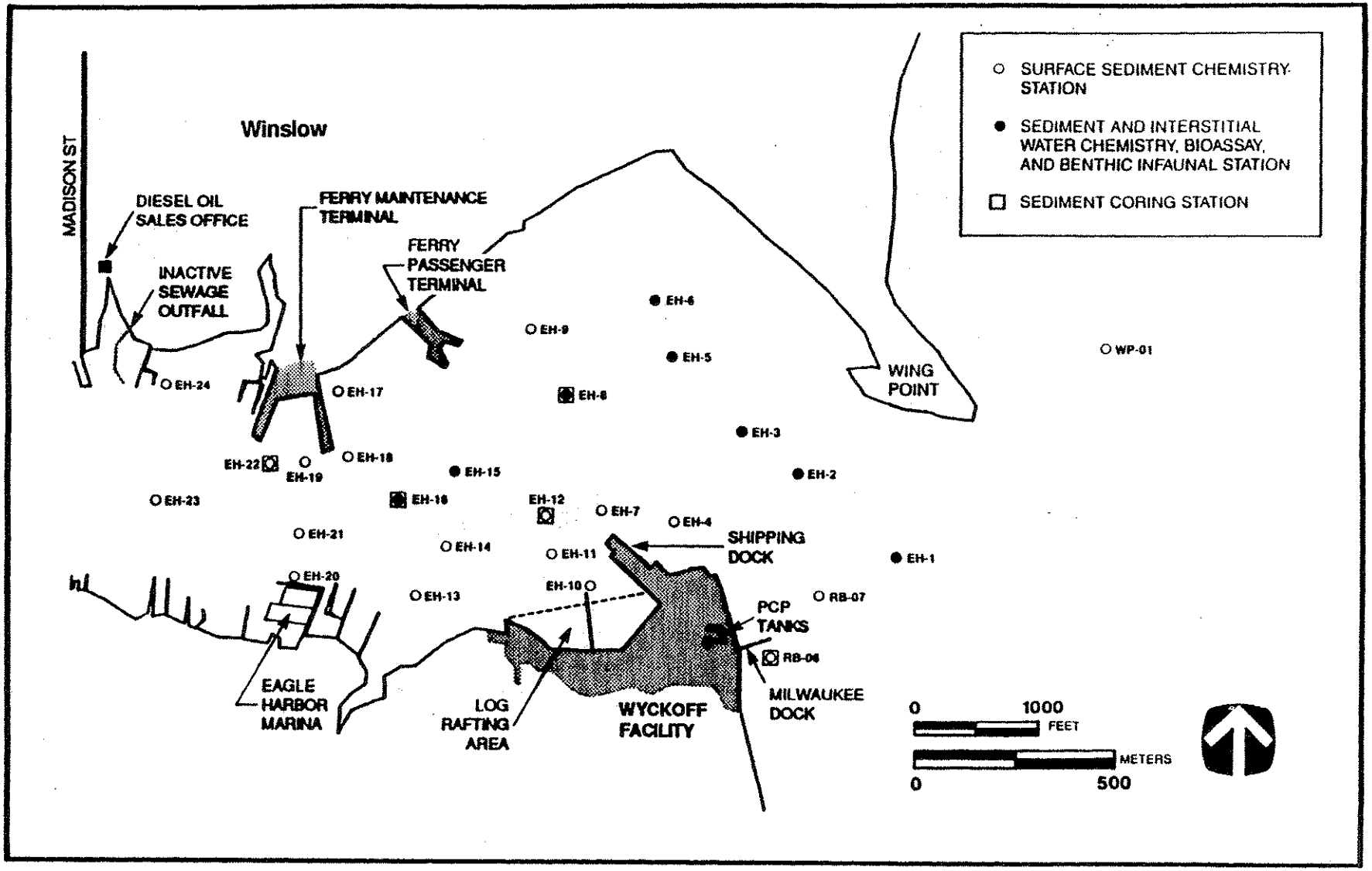


Figure B-19. Sample station locations in Eagle Harbor Preliminary Investigation. (Reference: Barrick et al. 1986).

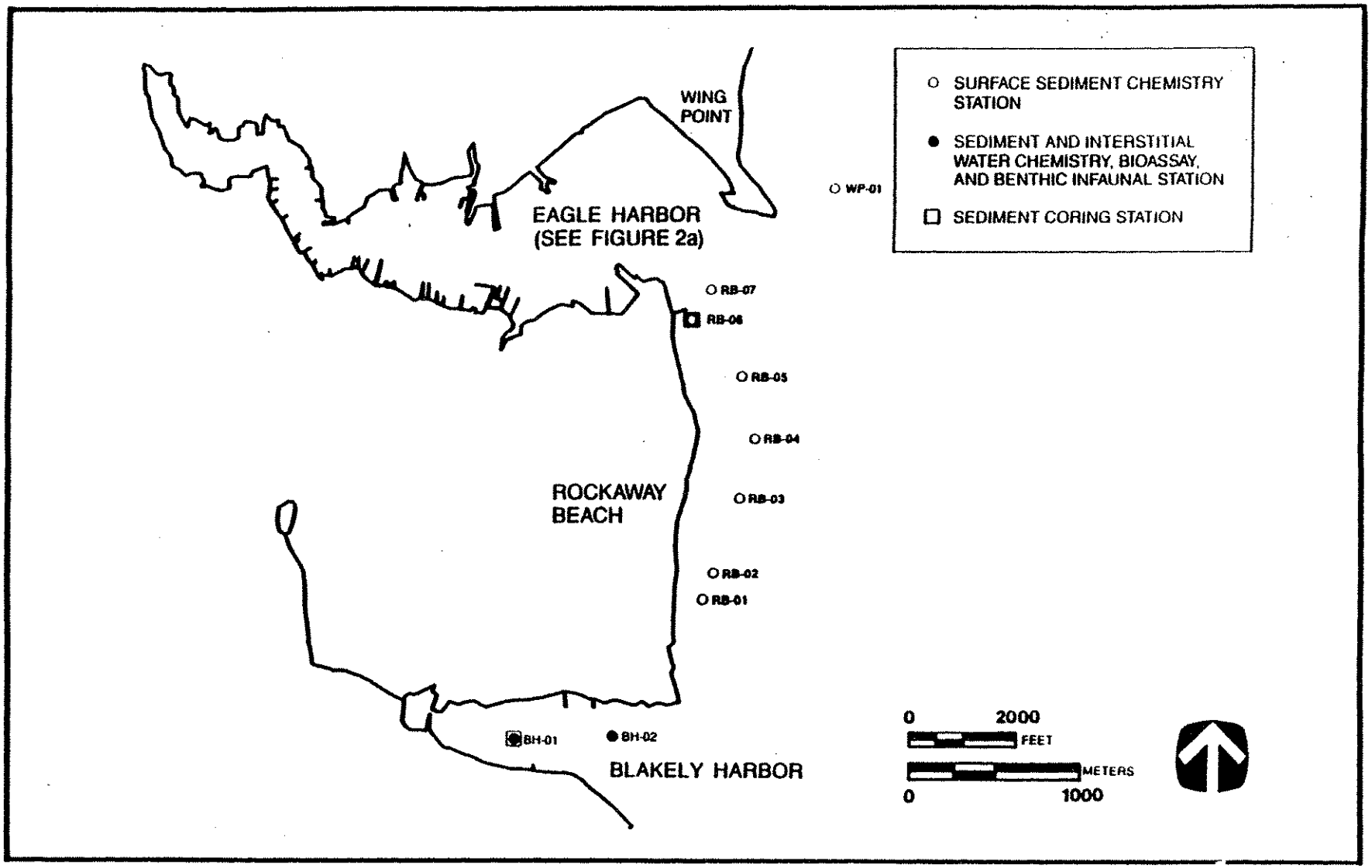


Figure B-20. Sample station locations in Blakely Harbor reference area. (Reference: Barrick et al. 1986).

APPENDIX C

SUMMARY OF OPTIONS AND RECOMMENDATIONS FOR TREATING ANOMALOUS DATA

CONTENTS

FIGURES

TABLES

SUMMARY OF OPTIONS AND RECOMMENDATIONS FOR TREATING ANOMALOUS DATA

Recommendations are presented in this appendix for procedures to "protect" the integrity of the sediment quality values database (SEDQUAL) from the inclusion of anomalous stations (from either a chemical or biological perspective). Integrity of the SEDQUAL database is important not only for developing AET, but also for validating any sediment quality value by comparing predictions of biological effects with observations stored in SEDQUAL. The following three issues are of primary concern:

- **Treatment of statistical power and Type I error in biological tests:** What is the preferred alternative for excluding inconclusive biological data from the sediment quality values database? (i.e., screening for biological tests with unusually low statistical power that result in erroneous classification of nonimpacted stations used to establish AET).

- **Data quality assurance for the sediment quality values database:** What qualified chemical data are acceptable for use in the database to establish or verify sediment quality values?

- **Representativeness of AET values:** What is the preferred alternative for excluding potentially anomalous chemical or biological data in the sediment quality values database? (e.g., high chemical values at stations that do not exhibit statistically significant biological effects).

Procedures that have been used in the past to address each of these issues are described in the following sections. Other options for addressing these issues in AET calculations are also discussed. A summary of recommended options based on this discussion is provided in Table C-1, including what options have been implemented in the current report for calculating AET. Note that not all of the recommendations could be implemented for every indicator.

Options that have been implemented to update Puget Sound AET in this project are also summarized in the METHODS section of the main text. A detailed discussion of how power analysis can be integrated into biological screening decisions is provided in Exhibit 1 of this appendix.

STATISTICAL TREATMENT OF BIOLOGICAL DATA

The statistical analysis of biological data in relation to reference data forms the basis for classification of stations as impacted or nonimpacted for use in developing AET. When evaluating biological data used to develop or validate sediment quality values, there are two primary concerns. First, appropriate reference area data should be used to test for biological effects at potentially impacted sites. Second, replicate data should have a low enough variance to ensure adequate statistical power to detect important differences between a potentially impacted site and the reference area.

Historically, study-specific reference conditions have been used to assess biological responses for AET calculations. Reference samples have typically been collected from

TABLE C-1. RECOMMENDED PROCEDURES FOR TREATMENT OF DATA USED TO DEVELOP SEDIMENT QUALITY VALUES

Issue	Recommendation	Status
Biological Data Evaluation		
Reference	Accept reference data that meets guidelines for mean values and standard deviation (e.g., mean amphipod mortality <25 percent and S.D. <20).	Implemented for amphipod bioassay
	Use one of reference data options described in text when full study-specific data set does not meet guidelines above.	Used partial reference data set for amphipod bioassay
Statistical comparisons	Test between-site differences using ANOVA with pairwise alpha of 0.05. Classify significant ($P<0.05$) differences from reference as "impacts", when the effects guidelines are exceeded (e.g., 25 percent mortality for amphipod bioassay, 50 percent depression for infauna).	Implemented for amphipod bioassay and infauna
	If apparent effect does not exceed the guideline, the site is classified as "nonimpacted."	Implemented for amphipod bioassay and infauna
Power analysis	Classify stations that are not significantly different from reference ($P>0.05$) with standard deviation (S.D.) less than guideline (e.g., S.D. <15 for amphipod bioassay) as "nonimpacted".	Implemented for amphipod bioassay
	For nonsignificant stations exceeding S.D. guideline:	Implemented for amphipod bioassay
	a) Classify as "nonimpacted" if variance does not exceed guideline corresponding to power = 0.6 (i.e., power is predicted as ≥ 0.6).	
	b) Classify as "inconclusive" if variance exceeds guideline corresponding to power = 0.6 (i.e., power is predicted as <0.6.	
Chemical Data Evaluation		
Chemical qualifiers	In calculating sediment quality values, exclude undetected values and data that are recovery-corrected by a factor >10.	Implemented, except for recovery-corrected data
AET representativeness	Accept nonimpacted stations for which there is confirming evidence within a concentration factor of 3 by at least one additional nonimpacted station.	Implemented
	Also require on a case-by-case basis, confirmation by at least one nonimpacted station at a different geographic location.	Not implemented

areas of Puget Sound that have relatively low chemical contamination and few biological impacts. Statistical analyses for classifying biological effects used to develop AET values have also focused on minimizing false positives as measured by Type I error (i.e., *alpha*, the probability of misclassifying a site as impacted when it is actually not significantly different from reference). A relatively low *alpha* level (e.g., pairwise t-tests at an effective *alpha*=0.001) was used historically in AET calculations (Beller et al. 1986). It is possible that anomalous classifications of stations may result because statistical power has not been explicitly considered during these analyses.

In this section, two options for treatment of anomalous biological data are reviewed. Second, a multistage procedure that combines these options for evaluating biological data is discussed. The procedures applied to evaluate biological data used to develop updated AET in this report are summarized at each stage.

Options for Treatment of Anomalous Data

The following two options can be used to identify biological observations that may be inadequate for developing or validating sediment quality values:

- **Screening-level Evaluation of Data:** Exclude all data that show unusually high variability compared with a guideline established from evaluation of data for the same biological variable in other Puget Sound studies.

- **Evaluation of Statistical Power:** Evaluate data variability relative to statistical power (i.e., **the capacity to detect significant effects**). The higher statistical power is, the more likely it is that a given difference from reference can be detected. Exclude all data where power of the statistical test is below a specified value (i.e., defined below as power <0.6).

As shown in the next section, the screening-level evaluation of data variability and the statistical power analysis could be applied in a complementary manner. The screening-level step can be used to ensure that the power analysis is applied efficiently (i.e., only to selected stations). The screening-level evaluation addresses only one variable (i.e., variance) that influences the likelihood of detecting biological effects. The evaluation of statistical power is more complex conceptually, but it addresses all key variables: data variability, reference values, the minimum detectable difference between the study site and reference site, and *alpha*.

Summary of Integrated Option for Evaluation of Biological Data

In this section, an integrated procedure is proposed for evaluation of biological data to support development of AET. The proposed procedure is illustrated using the amphipod bioassay as an example. The steps of the evaluation procedure would be essentially the same for all biological data types, although the numerical guidelines for evaluating data variability (e.g., standard deviation limits) may differ among data types.

The four stages of the proposed evaluation procedure are summarized below and in Figure C-1:

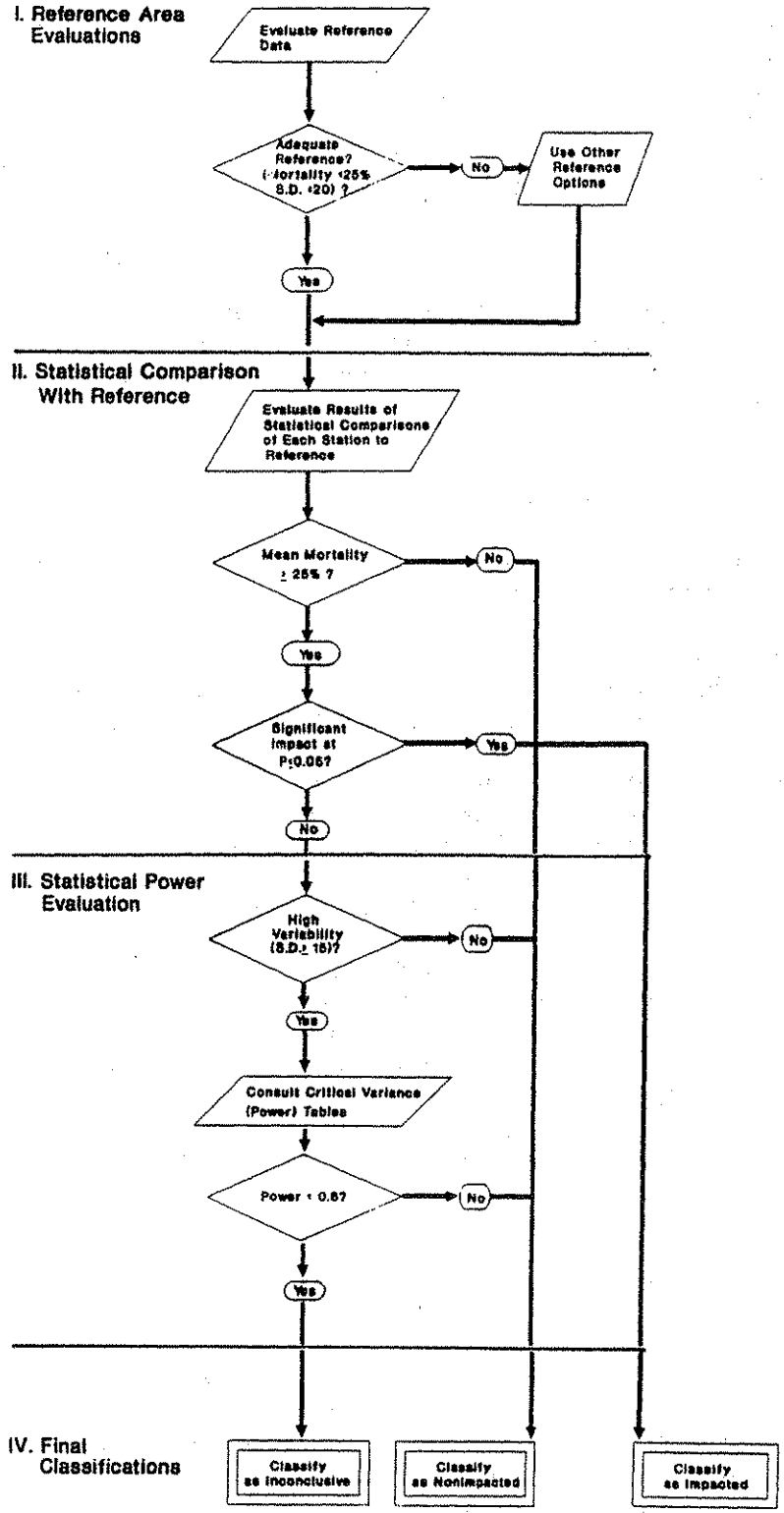


Figure C-1. Procedure for evaluation of statistical results of amphipod sediment bioassays.

Stage I: Evaluate reference-area data-Use of inadequate reference data (e.g., unusually high mean mortality or standard deviation) can reduce the ability to discriminate impacts. Historically, reference data in Puget Sound have not been screened for statistical adequacy. It is proposed that all reference data pass screening criteria outlined below before being used to assess impacts:

A. Compare the mean amphipod mortality and the standard deviation at the reference area with the guidelines of 25 percent and 20 percent, respectively.

B. If mean mortality, or the standard deviation exceeds the guideline, use alternative reference data in Stage II, which could include one of the following options:

   (1) Use only those study-specific reference stations that pass the criteria

   (2) Use historical data

   (3) Collect new data

   (4) Use standard reference data.

For the current project, guidelines in Step I-A were implemented for all amphipod bioassay data (see Table C-1). In Step I-B, reference stations that did not meet these guidelines were rejected (one reference station each from Carr Inlet, Sequim Bay, and Port Susan failed these guidelines; the remaining 75-80 percent of the reference stations in each area passed the guidelines). Those study-specific reference stations passing the criteria were used in all three cases.

Stage II: Perform statistical analyses to compare a potentially impacted site with reference conditions--Under certain study-specific circumstances (e.g., very low reference mortality, low variability at reference and potentially impacted stations), a station may be classified as impacted although the observed mortality level is within the range of reference conditions generally observed in Puget Sound. Historically, no criteria were used to distinguish such cases, but amphipod mortality less than 25 percent was rarely classified as statistically significant. It is proposed that mortality at any potentially impacted site exceed a minimum value before it is designated as impacted based on statistical comparisons. The following steps are proposed:

A. Perform pairwise comparison of each study site and the reference area using analysis of variance (ANOVA) program on available software.

B. From the ANOVA results, list all impacted sites (i.e., significantly different from reference, $P \leq 0.05$) and nonimpacted sites with low mean mortality ($P > 0.05$; <25 percent mortality). Proceed to Stage IV of evaluation procedure for these sites.

C. For nonimpacted sites with relatively high mean mortality ($P > 0.05$; ≥ 25 percent mortality), proceed to Stage III of evaluation procedure.

These Stage II guidelines were implemented in the current project for the amphipod bioassay (see Table C-1). Similar guidelines implemented for statistical comparisons for

other indicators (i.e., benthic infauna analyses) are summarized in Table C-1 and Section 2.4 in the main text.

Stage III: Evaluate statistical power--If statistical power is unusually low (see Exhibit I following this appendix), the ability to discriminate impacts can be reduced substantially. Historically, little consideration has been given to the power of statistical comparisons. The following steps are proposed to determine if potentially impacted stations with unusually high variability have been designated as nonimpacted because of inadequate statistical power:

A. List nonimpacted sites with standard deviation of mean mortality <15 and proceed to Stage IV for these sites.

B. For nonimpacted sites with standard deviation of mean mortality ≥15, compare the within-group variance from the ANOVA table with a critical threshold variance guideline corresponding to statistical power of 0.6.

C. List sites with variances that do not exceed the critical threshold variance as "nonimpacted". List results with variances that exceed critical threshold guidelines as "inconclusive".

This Stage III procedure was implemented in the current study for amphipod bioassay results.

Stage IV: Determine final classification of each station (i.e., impacted or nonimpacted)--Historically, test stations have been designated as impacted or nonimpacted without regard for statistical power. It is proposed that a new classification of "inconclusive" be used for cases in which power is considered inadequate. The final step, then is as follows:

A. Develop final classification of each station as "impacted", "nonimpacted", or "inconclusive results" based on the preceding evaluation. Inconclusive stations would not be used to develop sediment quality values.

This Stage IV procedure was implemented for amphipod bioassay results (see Table C-1). Stages of the proposed evaluation procedure and the rationale for guidelines are discussed below.

Rationale for Integrated Option

Stage I: Reference Area Evaluation--For the amphipod bioassay, the mean mortality and standard deviation of the mean at each reference area can influence the ability to discriminate a statistically significant increase in mortality at a potentially impacted station. An unusually high value of either parameter would tend to reduce this discriminatory ability. To guard against the influence of anomalously high values, reference conditions having a mean mortality ≥25 percent or a standard deviation ≥20 are considered inappropriate for statistical comparisons supporting AET (Figure C-1). The critical value for mean mortality is based on interlaboratory studies by Mearns et al. (1986), who concluded that values ≥25 percent could be reliably interpreted as representing sediment toxicity. Evaluation of mortality in the reference areas from the eight studies

in the SEDQUAL database showed that 3 of the 23 (13 percent) reference stations exceeded the critical value. The critical value for the standard deviation of mean mortality was derived by examining the values for the 23 reference stations from the eight studies in the SEDQUAL database. All but one value (i.e., 31) of the standard deviation ranged from 4.0 to 16.

As noted above, the ability of various studies to discriminate statistically significant responses using the amphipod mortality bioassay is influenced by the study-specific reference conditions used to evaluate the responses at potentially impacted stations. Several important characteristics of reference conditions that may influence study results include the two discussed earlier (i.e., mean mortality and standard deviation of the mean value), as well as the number of replicate analyses from each reference area. All three characteristics varied among the eight studies included in the SEDQUAL database. The number of stations sampled in each reference area ranged from 1 to 4, which, when pooled, yielded 4 to 20 replicate analyses. Mean mortality ranged from 8.0 to 22 percent (excluding stations with mean mortality $\geq$25 percent or standard deviation $\geq$20), and the standard deviation ranged from 5.0 to 12.

The influence of the differences in reference conditions among the eight studies in the SEDQUAL database is shown in Table C-2. The least-detectable elevated ($P \leq 0.05$) level of mortality ranges from 17 to 34 percent. Thus, statistical assessments of bioassay responses have probably been relatively consistent among studies for values of mortality less than 17 percent and greater than 34 percent, but inconsistent for cases that fall between these two values.

To eliminate the influence of study-specific reference conditions on the results of different studies, a set of standard reference conditions could be developed for comparison with all potentially impacted sites. For example, the standard conditions for the amphipod bioassay may be represented by the mean values of percent mortality (i.e., 13), standard deviation of the mean (i.e., 8.6), and number of replicates (i.e., 12) used in the eight studies currently included in the SEDQUAL database. Alternatively, a subset of the historical data could be selected randomly to represent the standard set of conditions.

In addition to improving among-study consistency, the use of standard reference conditions would eliminate the need to sample reference areas in future studies. If the use of study-specific reference conditions is continued, the standard reference conditions could be substituted for the study-specific conditions, when the latter data are found to be inappropriate for statistical comparisons (i.e., when mean mortality or the standard deviation of the mean are unusually high; Figure C-1).

A major drawback to using standard reference conditions is the failure to account for among-study differences in the sensitivity of the test organisms as a result of season and/or year of testing. Although both positive and negative control tests are conducted to ensure that organism sensitivity is within acceptable ranges, there may be enough latitude within these ranges to potentially result in substantial among-study differences in organism sensitivity. It therefore is recommended that the use of study-specific reference conditions be continued for AET development (this recommendation has been applied to all data sets now in SEDQUAL). The use of standard reference conditions is one of several alternatives when a study-specific reference data set is rejected.

TABLE C-2. LEAST DETECTABLE RESPONSES FOR THE AMPHIPOD MORTALITY BIOASSAY USING VARIOUS SETS OF REFERENCE CONDITIONS

Reference Area	Least Detectable Response (%)[a,b]
Carr Inlet[c]	17.0
Blakely Harbor[d]	18.0
Sequim Bay[e]	20.5
Sequim Bay[f]	30.0
Sequim Bay[g]	27.0
Sequim Bay[h]	21.5
Port Susan[i]	25.0
Port Susan[j]	34.0

[a] Units are percent mortality.

[b] Detectable response is a significant ($P \leq 0.05$) increase in percent mortality relative to reference conditions, assuming that the standard deviation at each impacted station equals 14 (i.e., a value representative of those observed at impacted stations in the SEDQUAL database).

[c] Barrick et al. (1985)

[d] Barrick et al. (1986)

[e] Battelle (1985)

[f] U.S. Navy (1985)

[g] Chan et al. (1985)

[h] Chan et al. (1986)

[i] Beller et al. (1988a)

[j] Pastorok et al. (1988)

Stage II: Statistical Comparisons--After determining the appropriate data set to represent reference conditions, statistical analyses are performed to test for significant differences between each potentially impacted site and the reference area. Each statistical test used in evaluating biological data involves evaluation of a null hypothesis as follows: "There is no significant difference in the biological variable of concern between the potentially contaminated site and the reference site". A two-sample analysis of variance (ANOVA, which is equivalent to a t-test) is used to test for differences between sites using a pairwise *alpha* of 0.05 (see Background section in Exhibit 1 for details). Before performing the ANOVA, an F_{max} test (or Cochran's C-test) should be used to test for heterogeneous variances. When the variances are heterogeneous, the data may be transformed (e.g., arcsine transformation for bioassay data) to stabilize the variances. If data transformation fails to correct heterogeneous variances, then use of a nonparametric test is recommended (e.g., Mann-Whitney U-test).

In establishing an *alpha* level for sediment quality values data, correction for multiple comparisons (e.g., Bonferroni's technique discussed by Miller 1981) is not recommended because the focus is on determining pairwise differences between individual stations, not on a bay-wide or sound-wide view. Therefore, consideration of experiment-wise error based on the total number of stations within the SEDQUAL database is not relevant or warranted. Moreover, correction for multiple comparisons in the past (e.g., Beller at al. 1986) was performed for each study individually (because reference conditions were independent among studies), leading to inconsistency in pairwise *alpha* among studies. For example, studies consisting of 100 stations would be evaluated using 1/10 the *alpha* level applied to a study consisting of 10 stations in the same area, which could lead to different results even if the data at a particular repeated station were identical.

Use of consistent 0.05 pairwise *alpha* for Type I error is environmentally protective compared to lower values of *alpha*. The resulting enhanced potential for false positives in making multiple comparisons to a reference data set is balanced by the necessity to achieve a sufficient power using established protocols (i.e., five replicates). By setting a consistent Type I error rate, the minimum power achieved should be greater than 0.6, except where variance is unusually high (Exhibit 1).

The outcomes of the statistical tests are evaluated next. If a significant difference from reference is found, the station is classified as impacted. When no significant difference is found, the data are evaluated further to ensure that acceptance of the null hypothesis (i.e., classifying the station as nonimpacted) is appropriate. For example, if the mean mortality in the amphipod bioassay is <25 percent, the station is classified as nonimpacted. This value was selected in this case because mortality less than 25 percent is not considered high enough to warrant concern. Also, Mearns et al. (1986) concluded that sediments could not be reliably classified as toxic if the mean mortality of amphipods was less than approximately 25 percent. Nevertheless, the statistical results should be reported in the event that a site requires further monitoring.

Stage III: Statistical Power Evaluation--Tests that show ≥25 percent mortality in the amphipod bioassay, but are not significantly different from reference are evaluated further (Figure C-1). For example, if the standard deviation is <15, which by best professional judgement indicates that excessive variability of the data is not a problem, the site is classified as nonimpacted. An observation of relatively high variability (standard deviation ≥15) leads to a complete evaluation of statistical power. The

standard deviation guideline of 15 was derived from examination of the distribution of standard deviations for all amphipod bioassay data in the SEDQUAL database (Figure C-2). Approximately 20 percent of the observed standard deviations (n=310) were greater than 15.

The rationale for analysis of statistical power and a detailed explanation of the procedure are presented in Exhibit 1. Briefly, the basic premise is that the variance of the mean biological response at a site needs to be below a critical variance threshold to ensure adequate statistical power. If observed variance does not exceed the critical threshold variance, the station is classified as nonimpacted (Figure C-1). A minimum statistical power of 0.6 (at an *alpha* of 0.05) is recommended for amphipod bioassay data to be used in developing AET (Exhibit 1). Statistical (ANOVA) results for stations with power <0.6 (i.e., variance greater than the critical value) are regarded as inconclusive, and the data point would be excluded from AET calculations. The variance associated with the ANOVA is represented by the mean square within-group error. This value can be easily obtained from the printout of the ANOVA results from any standard statistical software package. By comparing the mean square within-group error from the ANOVA table with this critical value, investigators will know whether to use the station in setting an AET value. This evaluation can be performed graphically by plotting *observed* variance vs. mean survival at a study site and comparing the position of the data point with a *critical variance curve* (Figure C-3; also see Exhibit 1). Alternatively, unusually high variance (and hence low statistical power) can be determined by consulting tabulated values of critical variance. Such tables were constructed to evaluate the statistical power of the amphipod bioassay tests for Elliott Bay and Everett Harbor.

Stage IV: Final Classification of Stations--Stations are classified as "impacted", "nonimpacted", or "inconclusive" based on the results of the previous evaluations.

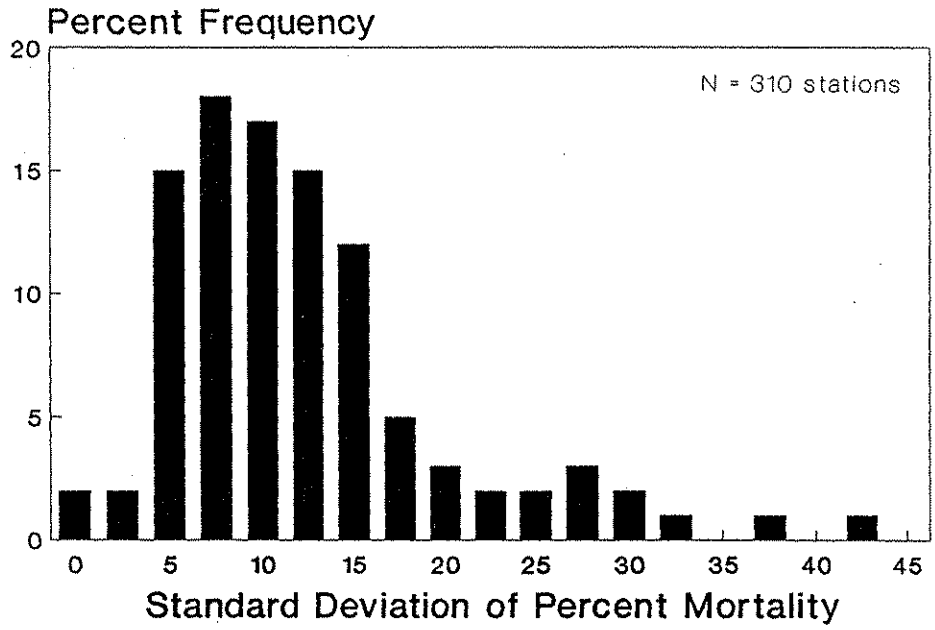
USE OF QUALIFIED CHEMICAL DATA

Following classification of impacted and nonimpacted stations, AET are determined from the associated chemical data at each station. AET historically have been developed using all detected data (i.e., excluding undetected data). Detected data that have been passed quality assurance review may still be qualified for one or more reasons. Qualified data (e.g., data qualified as estimates) may be less certain than data that have not been been qualified but the degree of uncertainty is typically undefined. Potential options for addressing the use of chemical data with qualifiers to develop AET include:

1. Exclude undetected data (i.e., data with a "B" or "U" qualifier) when establishing sediment quality values.

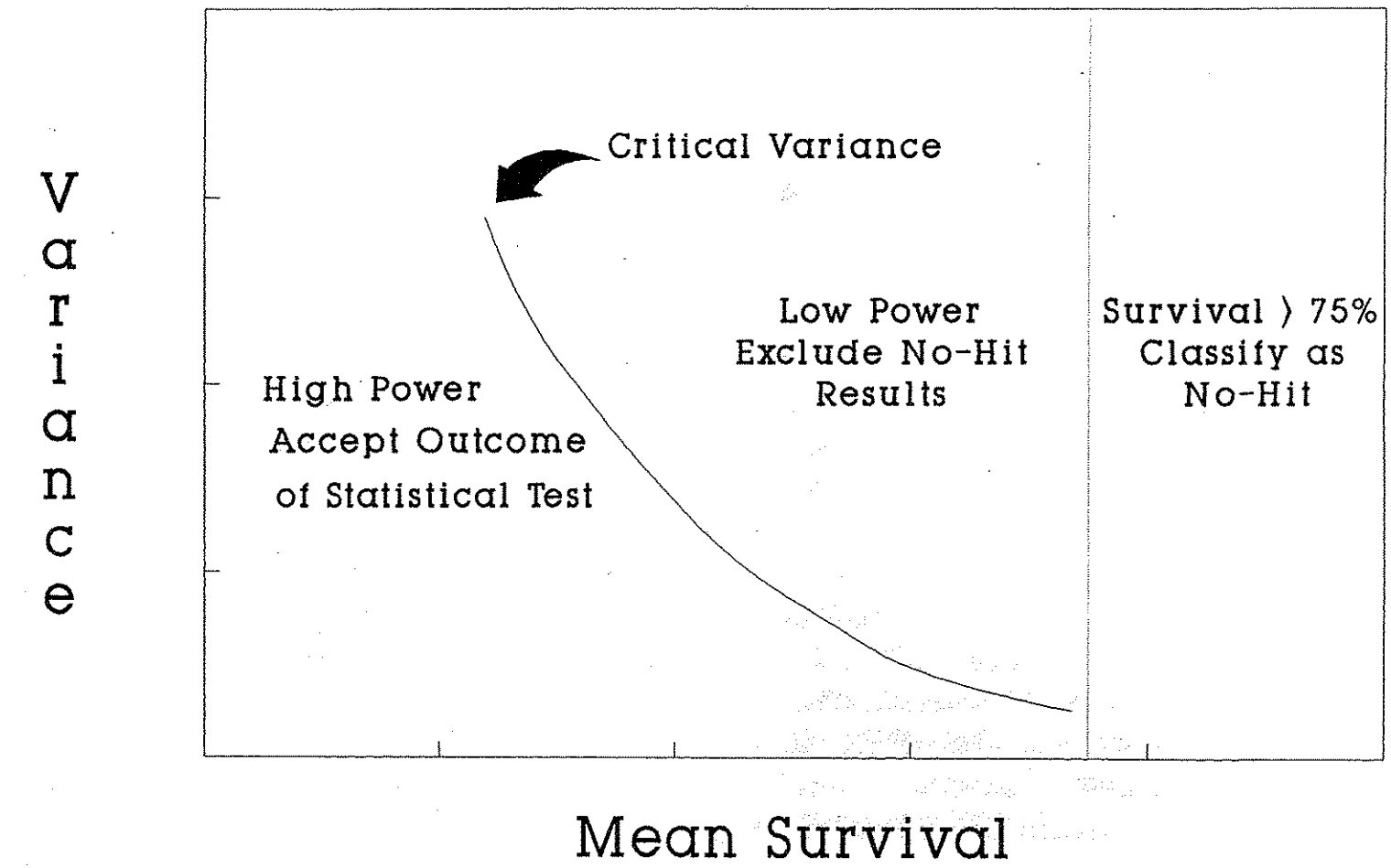
2. Exclude the above, and any data subjected to large recovery corrections (i.e., organic data with an "X" qualifier, indicating that greater than an order-of-magnitude recovery correction was applied based on surrogate recovery standards in the sample); note that for data that have not been recovery-corrected using the isotope dilution technique, a control limit of 50 to 75 percent recovery is required by PSEP protocols for data on organic compounds and metals.

3. Exclude the above, and any data that have been qualified as "estimates" (this option would have a major impact on use of many data sets, including use of any tentatively identified compounds which are by definition qualified with an "E").



Data Sources:
- Commencement Bay Remedial Invest. (n=56)
- Elliott Bay Action Program (n=106)
- Everett Harbor Action Program (n=34)
- Eagle Harbor Preliminary Invest. (n=10)
- 2 - COE Duwamish River Surveys (n=43)
- Eight Bays Reconnaissance Survey (n=54)
- Navy Everett Homeport Survey (n=7)

Figure C-2. Frequency distribution of standard deviations of percent mortality in amphipod bioassays for Puget Sound sediment quality values database.



Note: Critical variance curve is based on selected values for power and alpha (see text). Vertical line corresponds to 75% survival.

Figure C-3. Critical variance model.

Option 1 or 2 are recommended for development of sediment quality values. Option 2 addresses concerns about potentially setting AET too high (i.e., if large recovery corrections overestimate sediment concentrations), which is not considered in Option 1. Neither Option 1 nor Option 2 limit the appropriate use of data qualified as "estimates" such as Option 3 does. Data are qualified as "estimates" for a wide range of reasons in different programs; these reasons are considered as part of the quality assurance screening of data prior to entry into SEDQUAL.

AET in the current project were generated according to Option 1 (i.e., using only detected data but not excluding "X" or "E" qualifiers). The effect of implementing Option 2 by also excluding observations qualified by an "X" (recovery-corrected by greater than a factor of 10) was evaluated. The following AET would be affected by implementation of this option:

- 4-Methylphenol amphipod AET would decrease from 3,600 ug/kg dry weight to 2,600 ug/kg
- Acenaphthylene amphipod AET would decrease from 1,300 ug/kg dry weight to 560 ug/kg; benthic infaunal AET would decrease from 1,300 ug/kg to 1,200 ug/kg
- Pentachlorophenol amphipod AET would decrease from 360 ug/kg dry weight to 140 ug/kg.

These potential changes in AET would not affect the sensitivity of the 1988 benthic infauna AET and would affect the sensitivity of the 1988 amphipod bioassay AET only for acenaphthylene (an overall increase of roughly 6 percent in sensitivity for all amphipod bioassay surveys). This option is recommended for future analyses, subject to the approval of the Puget Sound Sediment Criteria Workgroup, if most extractable organic chemical analyses are to be performed without benefit of recovery correction of data using isotopically labelled surrogate compounds. Alternatively, the management of sediments could simply consider on a case-by-case basis the uncertainty implied by these results (e.g., that the 4-methylphenol amphipod AET of 3,600 ug/kg could be a lower value of 2,600 ug/kg).

REPRESENTATIVENESS OF AET VALUES

AET may change as additional data are added to the sediment quality values database. A change in an AET may result from addition of data to incorporate a more representative range of contaminant concentrations or from anomalous data. The degree of change and the ultimate stabilization of AET in relation to amount of data are discussed in the main report (see Results, Data Requirements to Derive Reliable AET). Anomalies that potentially lead to large increases in AET may be caused by nonrepresentative data, by exceptional chemical matrices (e.g., slag, coal), or unusual biological conditions (e.g., extremely tolerant species). Historically, no procedures have been used to identify chemically anomalous data. In this section, options for identifying and excluding (at least until additional data are available) anomalous data are evaluated, including the option implemented for the current project.

Options for addressing anomalous data can be classified into three general categories. First, anomalies may be implicitly treated by adopting a "safety factor" approach [e.g., using the lower limit of AET (Beller et al. 1986) or dividing the AET by a constant to establish sediment quality values]. Second, anomalies may be explicitly defined by

evaluating the magnitude of the difference in concentrations of a chemical between the station that sets the AET and the next nonimpacted station with a lower chemical concentration. Third, anomalies may be defined by modeling the distribution of nonimpacted and impacted stations. Options within each general category are listed below:

Safety Factor Approaches--

- **Lower limit of AET**: Lower limits of AET are evaluated in the PSDDA sediment quality values report (Beller et al. 1986). The lower limit of an AET is set protectively by assuming misclassification of two nonimpacted stations (Lower Limit 2), unless an administrative decision is made to assume just one misclassification (Lower Limit 1).

- **Fixed fraction of AET concentration**: Divide the AET by a constant safety factor to develop a sediment quality value.

Concentration Difference between Adjacent Nonimpacted Stations--

- **Fixed difference**: Accept only new nonimpacted stations that exceed existing Puget Sound AET by some factor (e.g., ≤ 3, thereby corresponding to a coefficient of variation of approximately 100 percent which is the largest control limit placed on chemical data by PSEP protocols).

- **Confirming Evidence (I)**: Accept nonimpacted stations for which there is confirming evidence within an arbitrary but consistent concentration factor (e.g., factor of 3) by at least one additional nonimpacted station

- **Confirming Evidence (II)**: Accept nonimpacted stations for which there is at least one additional nonimpacted station meeting the concentration criterion in (I) above at a different geographic location.

Modeling of Concentration Distribution--

- **Fixed percentile of distribution of nonimpacted stations**: Calculate the AET as the 95th percentile of the density distribution of the nonimpacted observations in the region just below the AET.

- **Runs test**: Determine the AET by performing a series of runs tests (i.e., a statistical test that is used to detect nonrandom sequences) on impacted and nonimpacted stations ordered according to a chemical concentration gradient. The AET would be identified by the station with the highest chemical concentration within the minimum set of stations that leads to a significant trend (i.e., nonrandom sequence of impacted and nonimpacted stations).

Confirming evidence is recommended as the best way to ensure that AET will not be set based on a localized condition that is not representative of Puget Sound conditions. Option (I) for confirming evidence has been implemented in the current project (see Table C-1). This option deals with anomalies explicitly, is simple to apply, and is consistent with the concept of the AET. The additional option (II) for requiring confirming evidence at a different geographic area would further reduce the likelihood of establishing

AET based on complete nonrepresentative matrices. However, determination of "different" geographic areas requires subjective evaluation and should probably be performed on a case-by-case basis as a policy decision using best professional judgement.

The fixed difference option ignores the potential for future confirmation of apparently nonrepresentative data and the value of periodically revising AET based on incorporation of additional data sets into SEDQUAL. The fixed difference option would also result in a potentially unwarranted exclusion of a large number of stations in the recent Elliott Bay and Everett Harbor PSEP surveys for certain chemicals. Other options (e.g., application of a runs test) do not explicitly identify anomalies or are limited by the need to model the density distribution of stations along a concentration gradient for individual contaminants.

Although safety-factor approaches are straightforward, their disadvantages outweigh this one advantage. Assuming safety-factor approaches would be applied consistently to every data set, the results would bear no relationship to data quality, amount of corroborative evidence (e.g., number of nonimpacted stations near the AET), or gaps in observations between the AET and the next highest chemical concentration for a nonimpacted station. Moreover, the safety-factor approach based on the fixed fraction of the AET would have a negligible effect when highly anomalous stations (e.g., those stations with concentrations >10 times the next highest nonimpacted station) set the AET. Yet, these cases are of the greatest concern.

Modeling of the distribution of single contaminants is only applicable to prediction of single-chemical effects on aquatic biota, which typically do not occur in most contaminated environments. Neither of the modeling alternatives (fixed percentage or runs test) account for the potentially random distribution of multiple causative agents in the environment. The underlying assumption of the AET concept is that no one chemical accounts for all biological effects observed in the environment. In particular, effects observed below the AET for a chemical are generally assumed to result from effects of other chemicals or environmental conditions; only above the AET are effects potentially associated with a chemical assumed to be discernable above these confounding environmental factors. Practical problems associated with chemical mixtures and matrix effects preclude application of any simple approach based on dose-response theory for single chemicals (e.g., such as may be applied in a controlled laboratory study for single chemicals).

APPENDIX C - EXHIBIT 1

EVALUATION OF STATISTICAL POWER IN BIOLOGICAL TESTS FOR AET DEVELOPMENT

BACKGROUND

Statistical comparisons of biological variables between contaminated and reference areas are critical in establishing AET. These tests are performed by evaluating a series of pairwise comparisons between potentially contaminated sites and reference sites using a two-sample analysis of variance (ANOVA). If no statistically significant difference in the mean value of a biological variable is found between a station and the reference value, it is categorized as a nonimpacted station. Since the AET is established by the nonimpacted station with the highest concentration of the chemical in question, the statistical errors associated with the ANOVA comparison for this station must be thoroughly understood and evaluated if investigators are to have confidence in the resulting AET value.

Each statistical test used in developing AET involves evaluation of a null hypothesis as follows: "There is no significant difference between the potentially contaminated site and the reference site". The degree of an investigator's confidence in the result of the statistical test depends on the magnitude of the two types of error associated with hypothesis testing (e.g., ANOVA). Type I error arises when the null hypothesis is incorrectly rejected, i.e. the null hypothesis is actually true but the investigator has rejected it. Type I error is expressed as the probability of making a mistake (*alpha*). Type II error occurs when the null hypothesis is accepted even though it is actually false and is similarly expressed as the probability of committing an error (*beta*). The power of a statistical test is defined as 1-*beta*, and represents the ability of a statistical test to correctly reject a false null hypothesis. For any given statistical test (mean, variance, and number of replicates held constant), as investigators reduce their *alpha* error, *beta* error is increased and statistical power decreased. Similarly, when *beta* error is reduced, statistical power and *alpha* error increase. It is thus desirable during any statistical test to reach a compromise between these two conflicting processes, i.e., maximize the statistical power (reduce Type II error) while minimizing the Type I error (*alpha*).

In the past, most investigators have controlled Type I error without considering or analyzing the statistical power of the test. Swartz et al (1985) examined the power of the standard amphipod bioassay protocol in the context of a t-test (which is equivalent to a two-sample ANOVA) and concluded that a difference of about 15 percent from laboratory controls could be detected at a power of 0.8 and a Type I error of 0.05. However, the analysis by Swartz et al. (1985) was potentially limited by the small range of test sediments and control conditions based on the native sediments for the amphipods. Evaluation of statistical power of biological tests for AET development requires consideration of the variability of the data sets used to develop AET and the influence of reference area results.

OBJECTIVES

The purpose of this analysis is to present an approach for screening the biological data used in setting AET values. The specific objectives addressed in this appendix are to:

- Describe a technique that allows investigators to evaluate the power of the statistical test used in defining nonimpacted stations by comparing the variance of their data with a critical level of variance corresponding to a selected level of statistical power
- Evaluate the effect of choice of *alpha* level on the capacity of a data set to meet minimum requirements for statistical power (i.e., percentage of stations that exhibit power greater than a selected minimum value)
- Apply the model for evaluating statistical power to amphipod bioassay data from Elliott Bay and Everett Harbor
- Recommend procedures for evaluating biological data used to develop AET to ensure adequate statistical power and confidence in nonimpacted observations.

Using the recommended approach, a station would not be used in setting the AET limit if the statistical power is judged to be inappropriately low for the selected *alpha* level. When this approach is implemented, investigators will have confidence that the decision to use a station in setting the AET limit for a particular chemical is based on known and acceptable statistical error rates.

CRITICAL VARIANCE MODEL

The power associated with an ANOVA test can be computed as a function of the following variables:

- Unexplained sample variance
- Minimal detectable difference between the two means
- *Alpha* level
- Number of stations
- Number of replicate observations at each station.

Using the procedures described by Scheffe (1959) and further elaborated by Cohen (1977), it is possible to calculate any one of these variables given values for the other five. Thus, given the desired statistical power (1-*beta*), the Type I error rate (*alpha*), and minimal detectable difference along with the actual number of stations and replicate observations, a *critical sample variance* can be computed. This critical sample variance can be used to represent the maximum tolerable variance that will meet the stated objectives for statistical power and *alpha* level.

When computing AET, the basic premise is that the variance of the mean biological response at a site needs to be below this critical variance threshold to ensure adequate power. The variance associated with the ANOVA is represented by the mean square within-group error. This value can be easily obtained from the printout of the ANOVA results from any standard statistical software package. By comparing the mean square within-group error from the ANOVA table with this critical value, investigators will know whether to use the station in setting an AET value.

The approach to evaluation of statistical power is illustrated below using amphipod bioassay data. A similar approach can be applied to other bioassays (e.g., oyster larvae) or benthic infauna. The critical variance model was presented graphically in Figure C-3 (Appendix C). The curve in Figure C-3 represents the plot of critical variance values (mean square within-group error) in relation to values of *mean percent survival* measured in the amphipod bioassays performed on sediments from potentially contaminated sites. The critical variance at a given value of percent survival is the variance at which statistical power is equal to the specified value at a selected *alpha* value (e.g., Figure C-3). As percent survival increases, the critical variance value decreases because precision of the data must be higher to detect smaller differences between a potentially contaminated site and the reference site (with high survival).

The example critical variance curve in Figure C-3 was derived by the calculation of statistical power for ANOVA based on Scheffe (1959) and Cohen (1977). A series of critical variance curves for specified values of power, *alpha*, and reference conditions could be easily derived for use with future data sets. As shown in later examples, observed values of mean square within-group error can be plotted in relation to observed mean survival values for potentially contaminated sites to evaluate observed variance relative to a specific critical variance curve. In the region above the critical variance curve, the power is less than the specified value. In the region below the curve, the power is greater than the specified value. Given selected values of power and *alpha*, data for the nonimpacted stations with variances above the critical variance curve would be excluded from AET calculations due to inadequate power.

In practice, the power criterion is applied only to nonimpacted stations. A theoretical calculation of low power is irrelevant to the case of an observed significant impact. Conversely, some observations of no impact should be accepted without evaluating power. For example, data would be subjected to an evaluation of power only if survival is less than 75 percent. The value of 75 percent survival was selected in this case because mortality less than 25 percent is not considered high enough to warrant concern. Mearns et al. (1986) concluded that sediments could not be reliably classified as toxic if the mean mortality of amphipods was less than approximately 25 percent.

APPLICATION OF THE MODEL TO AMPHIPOD BIOASSAY DATA FROM ELLIOTT BAY AND EVERETT HARBOR

The critical variance model was applied to amphipod bioassay data from Elliott Bay and Everett Harbor to illustrate the model and to evaluate the influence of the choice of values for *alpha* and power.

Methods

Amphipod (*Rhepoxynius abronius*) bioassay data were obtained from the 1985 survey of Elliott Bay (n=102 stations) and 1986 survey of Everett Harbor (n=29 stations)

for development of urban bay action plans as part of the Puget Sound Estuary Program [(Beller et al. (1986); Pastorok et al. 1988)]. ANOVA were performed on data files generated from SEDQUAL using SPSS. To examine the effect of data transformation, the ANOVA were performed twice: once using untransformed data and once using data modified by application of the arcsine transformation. The rationale for arcsine transformation of proportion data is provided by Sokal and Rohlf (1981).

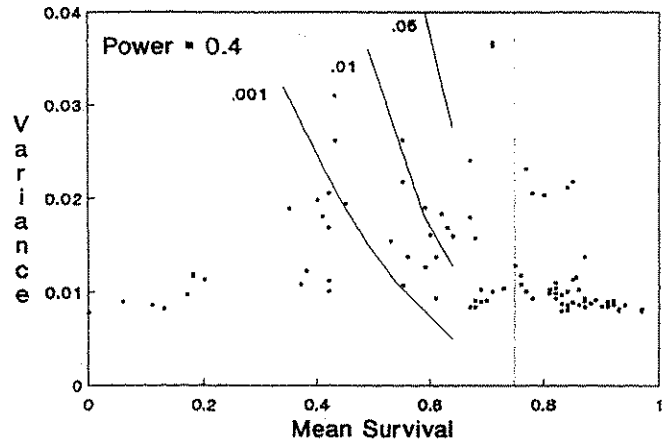
The amphipod bioassay data for Elliott Bay and Everett Harbor represent data sets where sample variability was high at certain stations and where survivorship in the reference area (Port Susan) was low to moderate relative to other reference areas in Puget Sound (Pastorok et al. 1988). Reference stations for these two surveys exhibited mean amphipod survival of 84 percent (1985 data; n=20) and 78 percent (1986 data; n=10). Because these data sets contain some stations with high variance (and therefore low statistical power), they illustrate the need for evaluating statistical power using the critical variance approach.

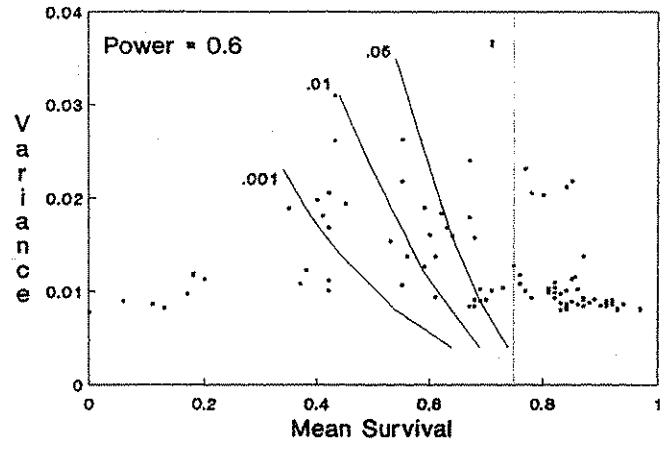
The techniques of Scheffe (1959) and Cohen (1977) were used to calculate critical variance values for a variety of conditions. Calculations were based on three *alpha* levels (.05, .01, .001), three power levels (0.8, 0.6, 0.4), two stations with five replicates per station, and a minimal detectable difference ranging from 20 percent to 50 percent of the reference station mean. Results were plotted separately for untransformed and arcsine transformed cases as curves of critical variance (mean square within-group) versus mean survival for the potentially impacted station (as in Figure C-3). The within-group mean square variances from the ANOVA performed on arcsine transformed data were compared with the corresponding transformed critical variances. Since the actual numbers of replicates at reference stations were 10 (Everett Harbor) and 20 (Elliott Bay), not five as was used in the calculations of critical variance, the power resulting from application of the model represents a minimal, and hence conservative, value.

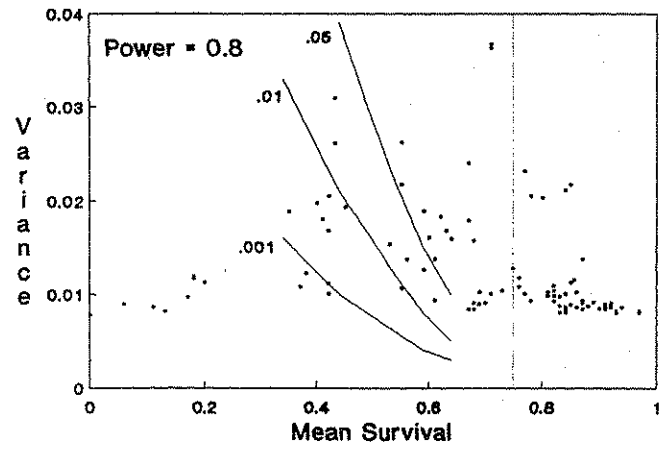
Results

Figure C-4 shows the critical variance curves for three power levels and three alpha values based on 1985 Port Susan reference data (mean amphipod survival = 0.84). The corresponding data points for the variance of mean survival based on the ANOVA for each station in Elliott Bay are also shown in Figure C-4. As the *alpha* level decreases for a given power, the lines of critical variance shift toward the left. The result is that more stations are classified as having insufficient power and are potentially eliminated from the list of nonimpacted stations used to generate AET. As the *alpha* level increases (e.g., .001 to 0.05), these curves shift toward the right and fewer stations are excluded. Similar shifts are seen when *alpha* is held constant but the desired power varies. At low power values (e.g., 0.40), fewer stations would potentially be eliminated from AET calculations than at higher values (e.g., 0.80). The vertical line in Figure C-4 corresponds to 75 percent survival (25 percent mortality). Stations to the right of the vertical line would be classified as nonimpacted.

Figure C-5 shows the critical variance curves and observed values of variance for both Elliott Bay and Everett Harbor using both untransformed and arcsine transformed data. The critical variance curves are presented for different *alpha* levels at a constant power of 0.60. Using the arcsine transformation resulted in a decrease in power (or alternatively, rejection of a greater number of nonimpacted stations for a given Type I error and power; Figure C-5). Because of the uncertain effects of transformation, it is recommended that data be transformed on a station by station basis to correct heterogeneous variances (cf. U.S. EPA and U.S. Corps of Engineers 1977; Mearns et al. 1986).



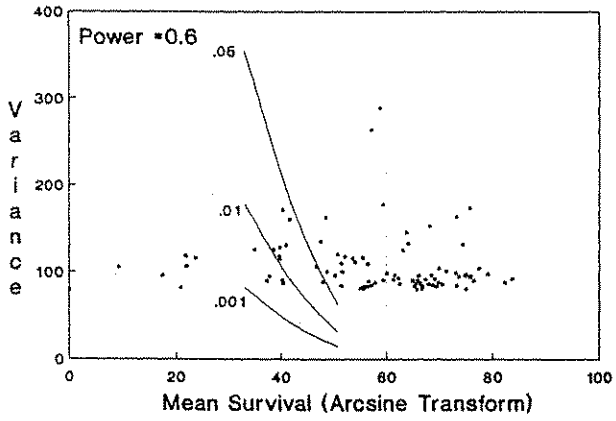


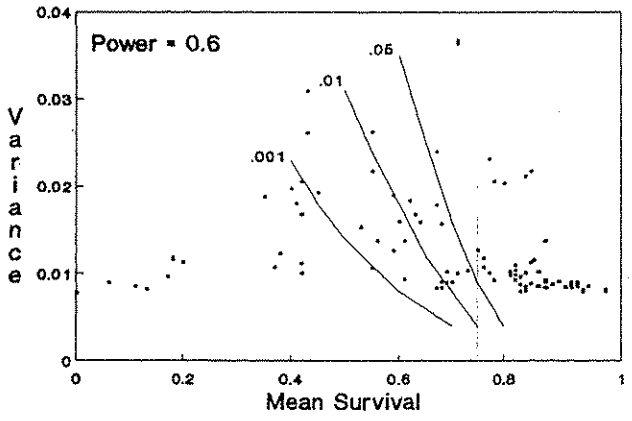


Note: vertical line corresponds to 75 percent survival.

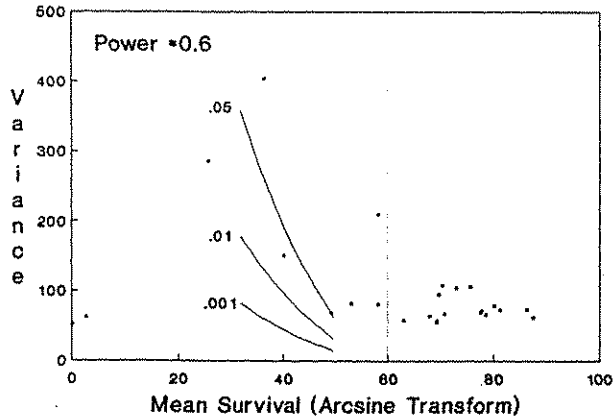
Figure C-4. Critical variance curves applied to Elliott Bay data for power of 0.4, 0.6, and 0.8 and *alpha* of 0.001, 0.01, and 0.05.

Elliott Bay





Everett Harbor



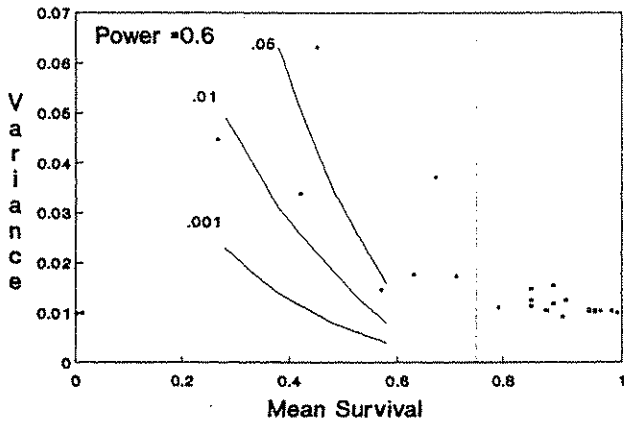


Figure C-5. Critical variance curves applied to Elliott Bay and Everett Harbor data for power of 0.6 and *alpha* of 0.001, 0.01, and 0.05 using untransformed and arcsine transformed data.

Of the 102 stations in Elliott Bay, 32 exhibited heterogeneous variances. Of these, data transformation resulted in higher variability (as measured by the square root of within-group mean square variance divided by the grand mean) and larger percent difference between means (as measured by the difference between means divided by the grand mean) in 6 cases, higher variability and lower difference between means in 13 cases, and lower variability and lower difference between means in 13 cases. Based on examination of the magnitudes of the changes in variability and in differences between means, the latter was clearly a more important factor than the change in variability. In most cases, variability of the transformed data was approximately the same as that of the untransformed data.

Based on the preceding analysis, a power of 0.6 and a Type I error of 0.05 appears appropriate for bioassay data analysis. Approximately 14 of the 134 stations (10 percent) in the combined Elliott Bay and Everett Harbor database fall in the region above the critical variance curve based on a power of 0.06 and an *alpha* of 0.05. Significant impacts were found at seven of these 14 stations. At the other seven stations, the results would be interpreted as inconclusive (i.e., a finding of "no impact" would be excluded from the SEDQUAL database due to inadequate power). Using a higher power or a lower Type I error to set critical variance levels would be inappropriate because of the large overlap between the rejection region and the bulk of the data points. The analysis also shows that the variance associated with values of mean survival less than about 45-50 percent is irrelevant as long as the *alpha* criterion is $\geq$0.01 and the power criterion is $\leq$0.6, (i.e., power would always be adequate due to a large difference between the study site mean and the reference mean). This conclusion is based on the assumption that reference survival is always greater than or equal to 75 percent.

RECOMMENDED PROCEDURES FOR EVALUATION OF STATISTICAL POWER

In the process of updating AET, the following procedures are recommended for the analysis of statistical power of amphipod bioassay data:

- **Set Type I error equal to 0.05 for all pairwise comparisons for all data sets.** Use of 0.05 for Type I error is environmentally protective compared to lower values. Although this may result in some false positives experimentwise, it is necessary to achieve a minimum power of 0.6 (especially when data are variable or survival in reference-area sediments is <85 percent). Lowering power further would be extremely difficult to justify. Higher power will be achieved where data are less variable than the data sets examined here or where reference values are $\geq$85 percent survival.

- **Evaluate bioassay data using the variance criterion based on a power of 0.6 and a Type I error of 0.05.** This evaluation should be completed for Elliott Bay and Everett Harbor before updating AET. Because the critical variance model was developed using these data, application of the model requires little additional effort. For other bioassay data sets, it requires performing ANOVA for all nonimpacted stations. Observations for nonimpacted stations with survival of 75 percent or greater and observations for impacted stations should be accepted regardless of power (i.e., the evaluation should not be performed for these stations). Nonimpacted stations with inadequate power should not be used to set AET.

- **Set AET using the resulting data and evaluate the results.**

Correction for multiple comparisons is not recommended because the focus of the present studies is on establishing pairwise differences between stations, not on a bay-wide or sound-wide view. Therefore, experimentwise error is not of great concern.

Eventually, similar data-screening procedures should be implemented for other biological tests. Power analysis of other data is beyond the scope of the present work assignment. Although the critical variance curves are independent of the study site (except to the extent that a station is evaluated relative to a specific critical variance curve associated with a specific mean survival in a reference data set), the distribution of points on the graph depends on the residual variance of the combined study site and reference-area data set. Because such procedures can not be implemented for other biological data within the present scope, interim procedures should be adopted. (For a complete discussion of procedures for evaluation of biological data used to develop AET, refer to the Methods section of the main text of this report).

APPENDIX D

LOCATION BY REGION OF STATIONS WITH PREDICTED BIOLOGICAL EFFECTS

FIGURES

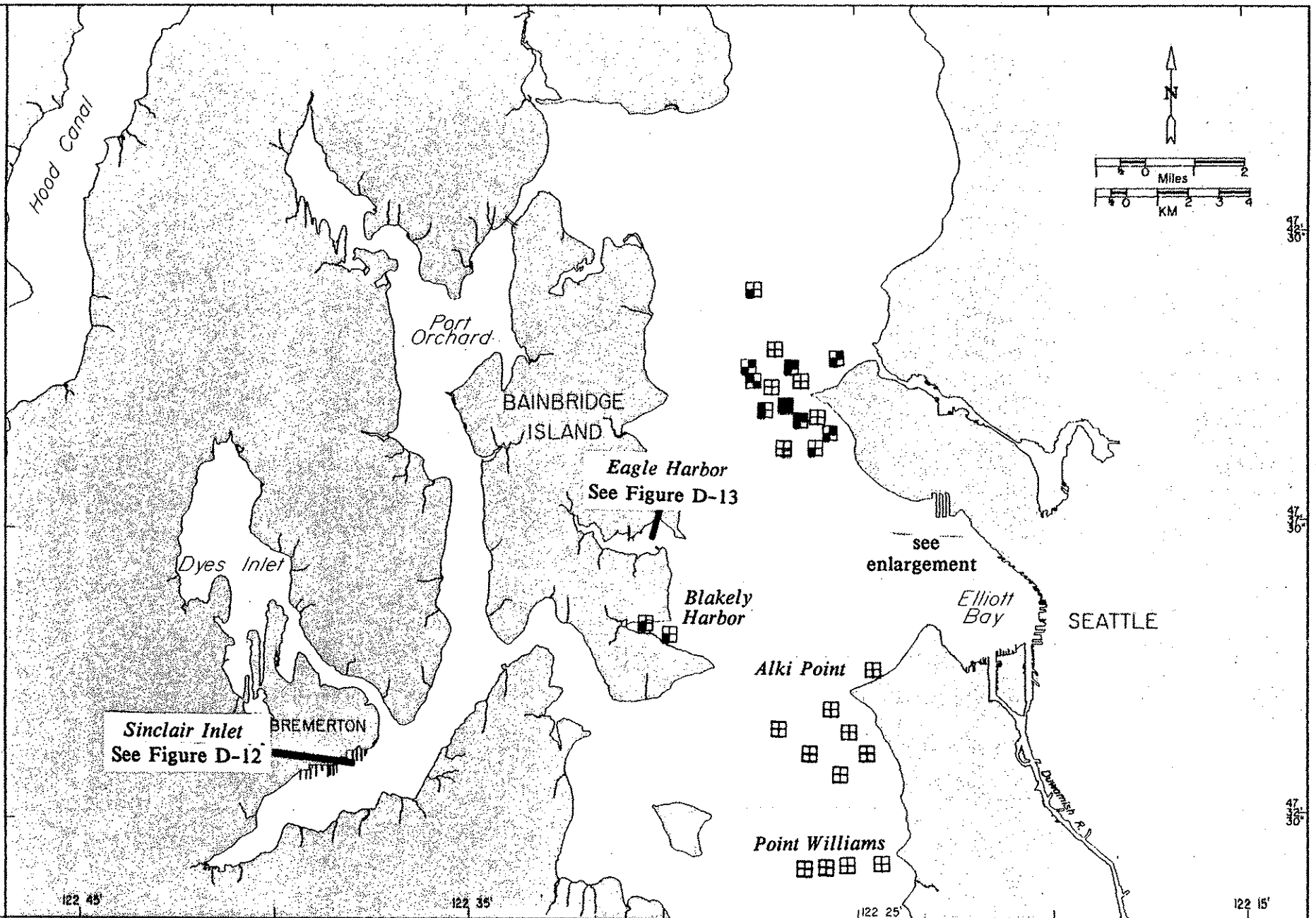


Figure D-1. Locations of predicted biological effects in central Puget Sound (predicted using 1988 Puget Sound AET).

Cartography by D.R.Systems,Inc.

Map compiled by Evans-Hamilton,Inc.

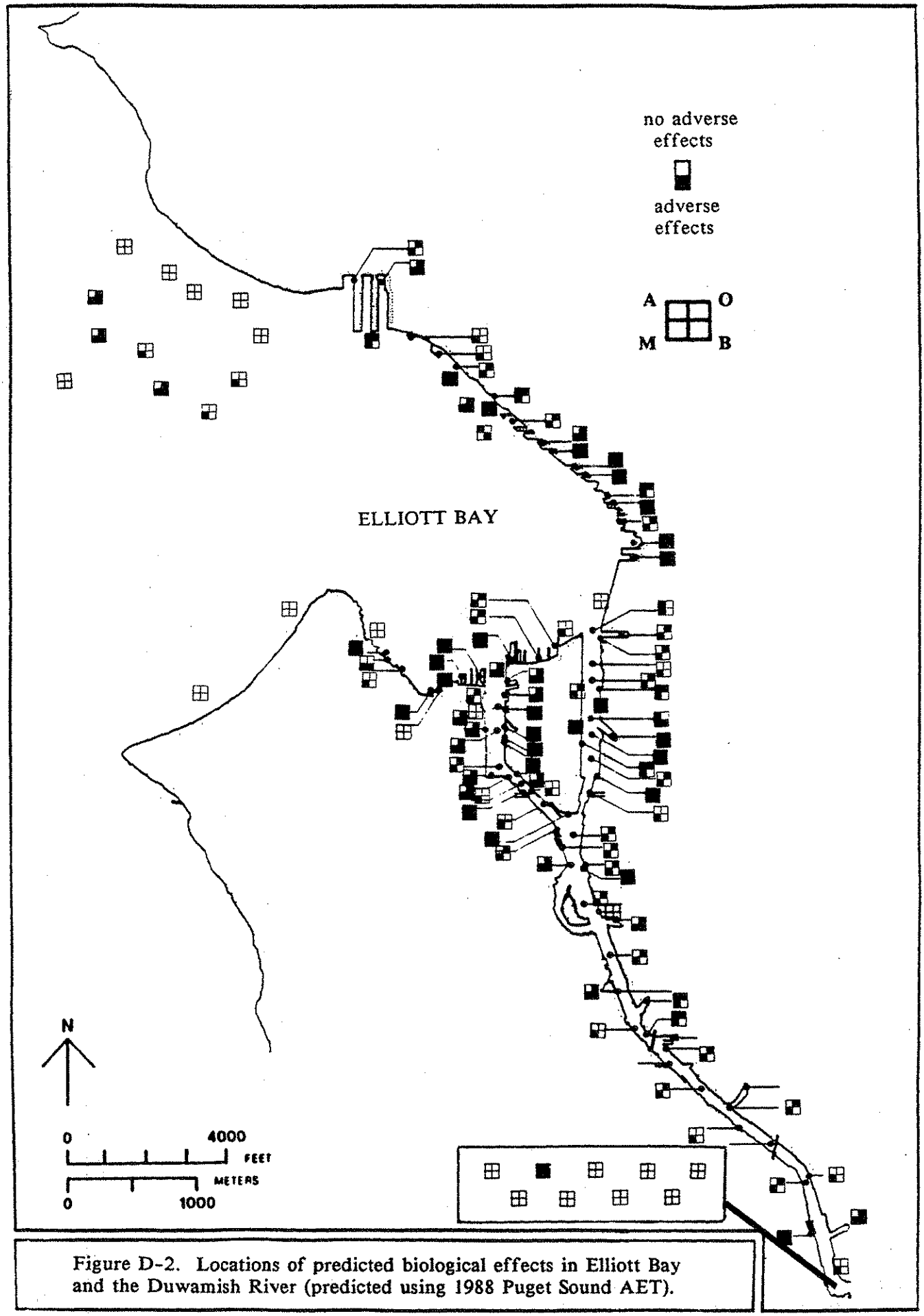


Figure D-2. Locations of predicted biological effects in Elliott Bay and the Duwamish River (predicted using 1988 Puget Sound AET).

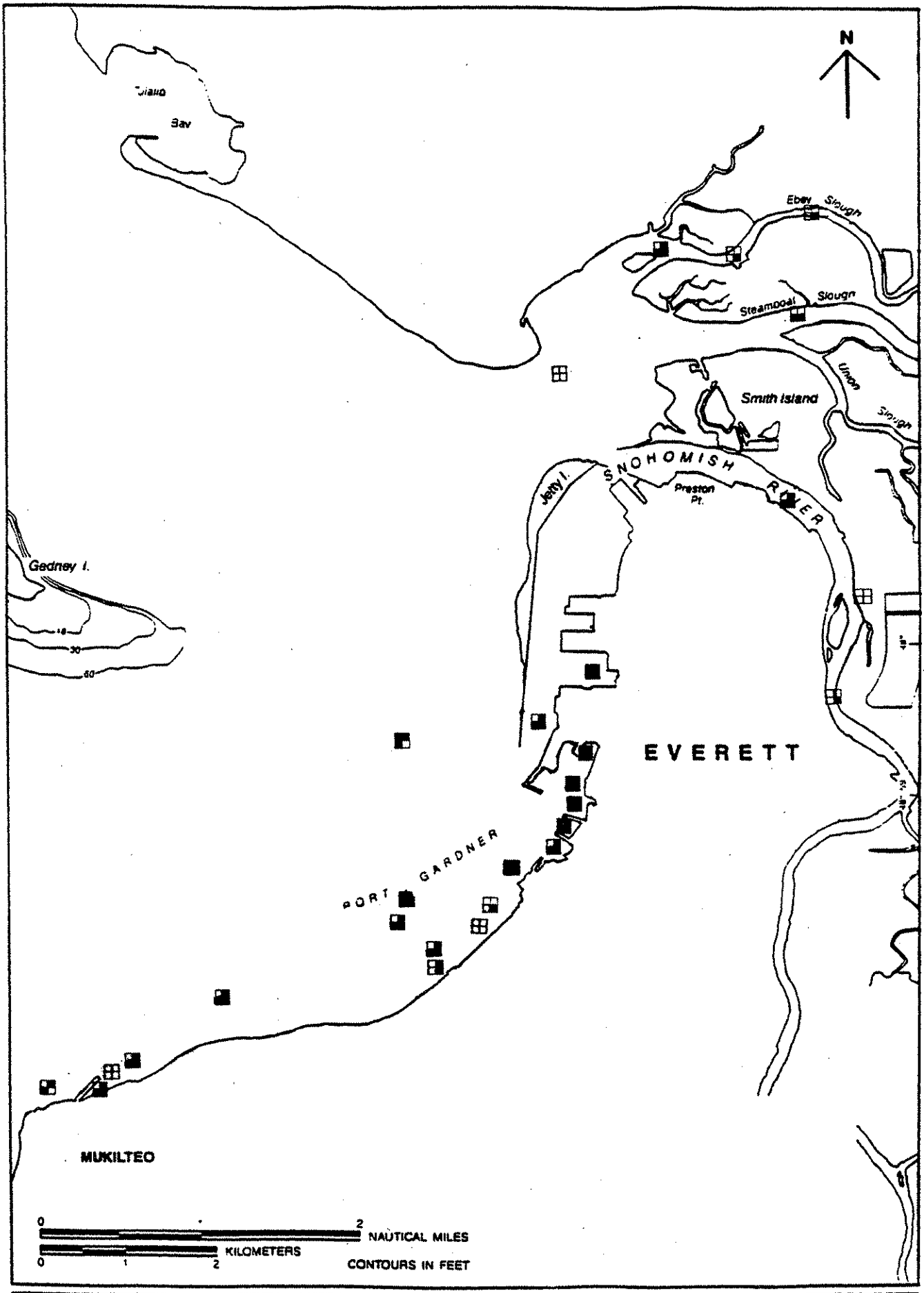


Figure D-3. Locations of predicted biological effects in Everett Harbor (predicted using 1988 Puget Sound AET).

no adverse effects
adverse effects

A O M B

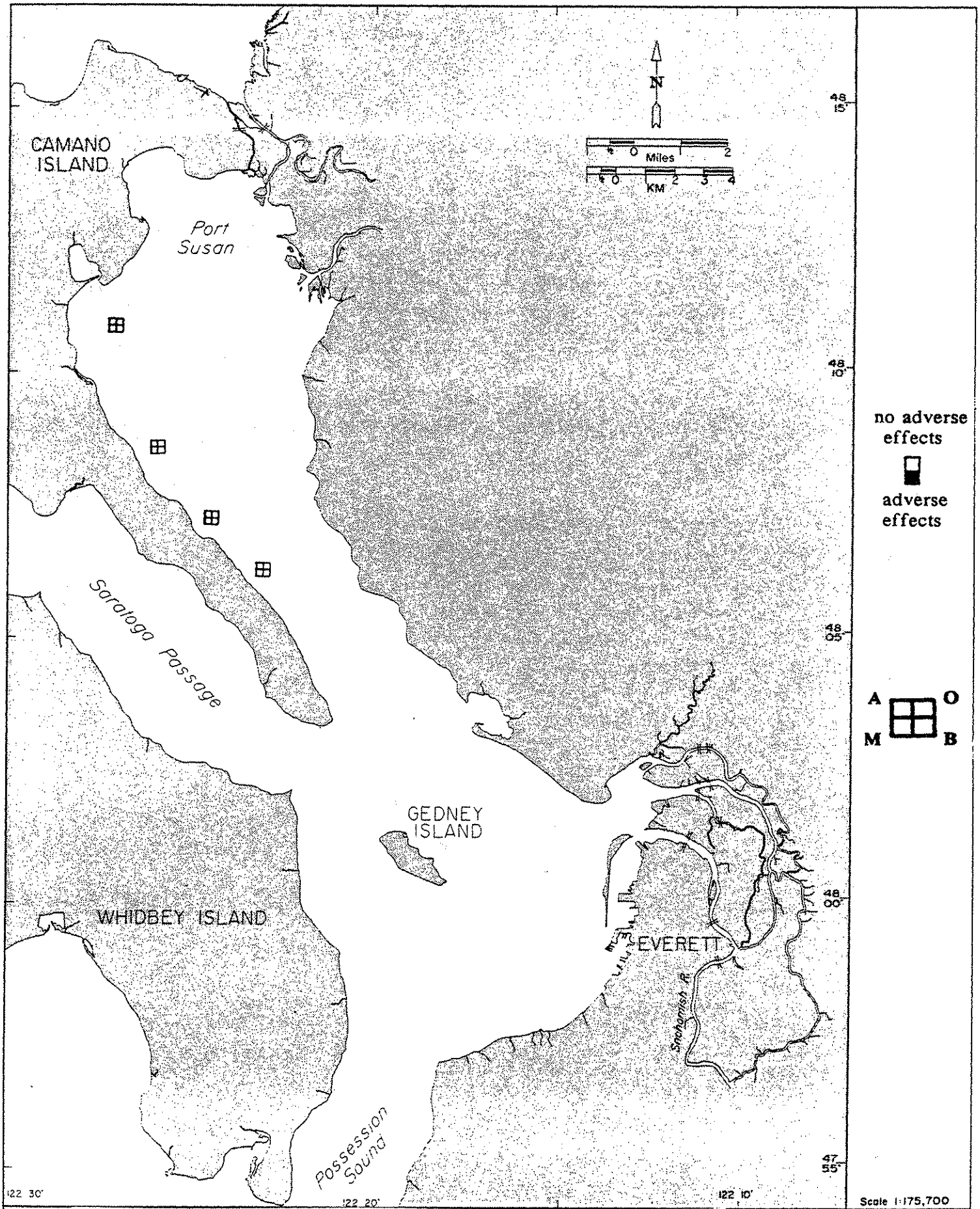


Figure D-4. Locations of predicted biological effects in Port Susan (predicted using 1988 Puget Sound AET).

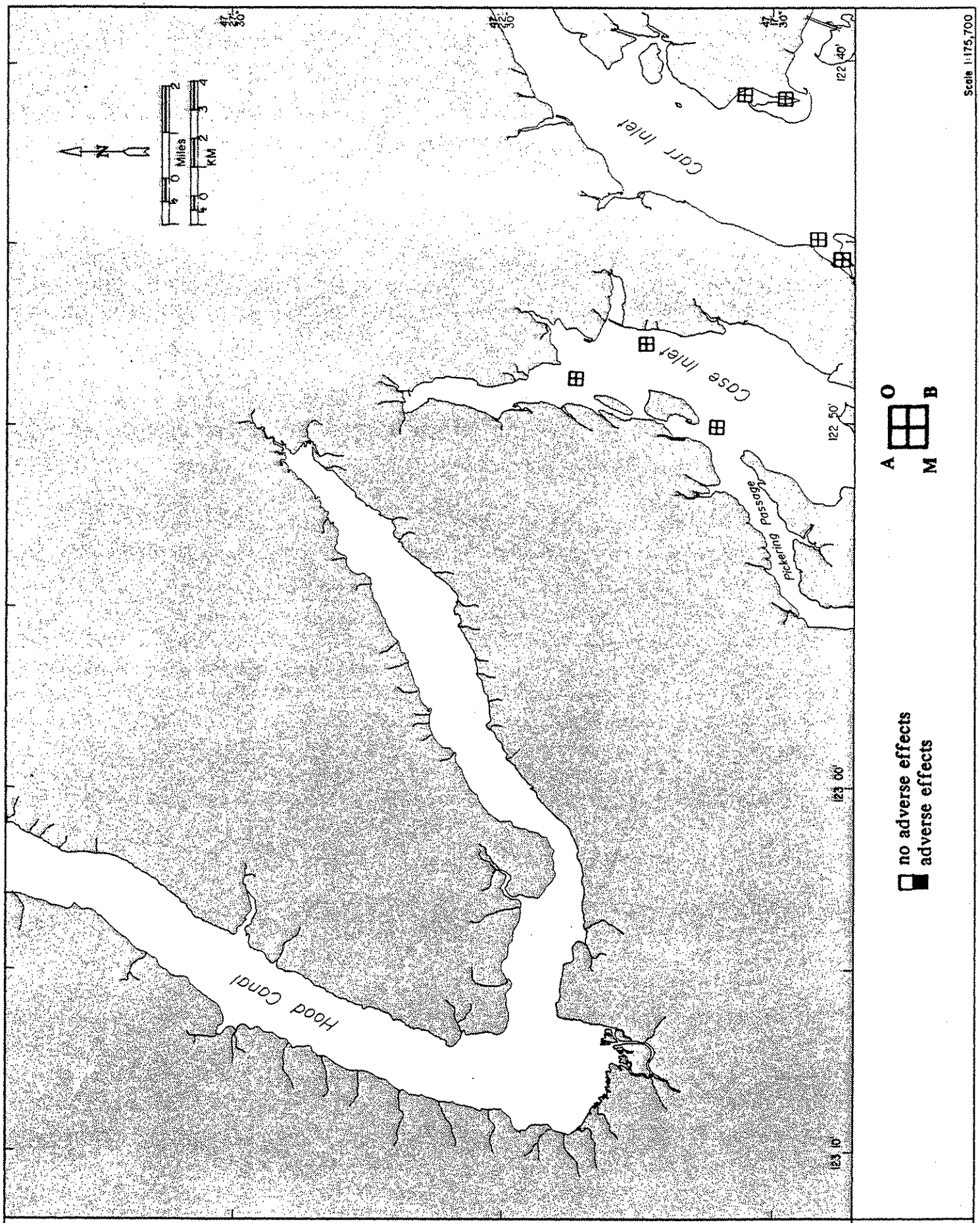


Figure D-5. Locations of predicted biological effects in Carr Inlet and Case Inlet (predicted using 1988 Puget Sound AET).

Figure D-6. Locations of predicted biological effects in lower Case Inlet (predicted using 1988 Puget Sound AET).

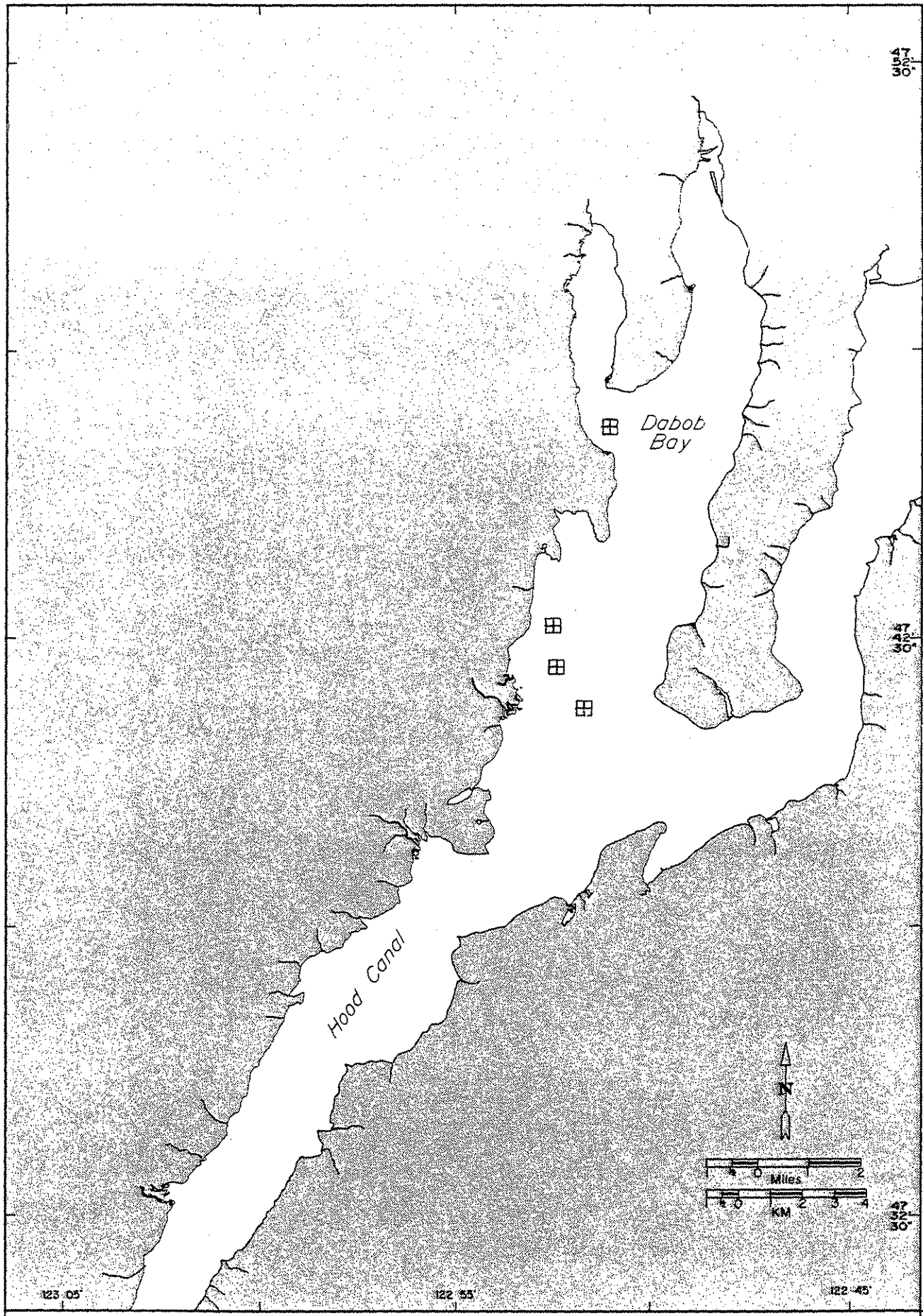


Figure D-7. Locations of predicted biological effects in Dabob Bay (predicted using 1988 Puget Sound AET).

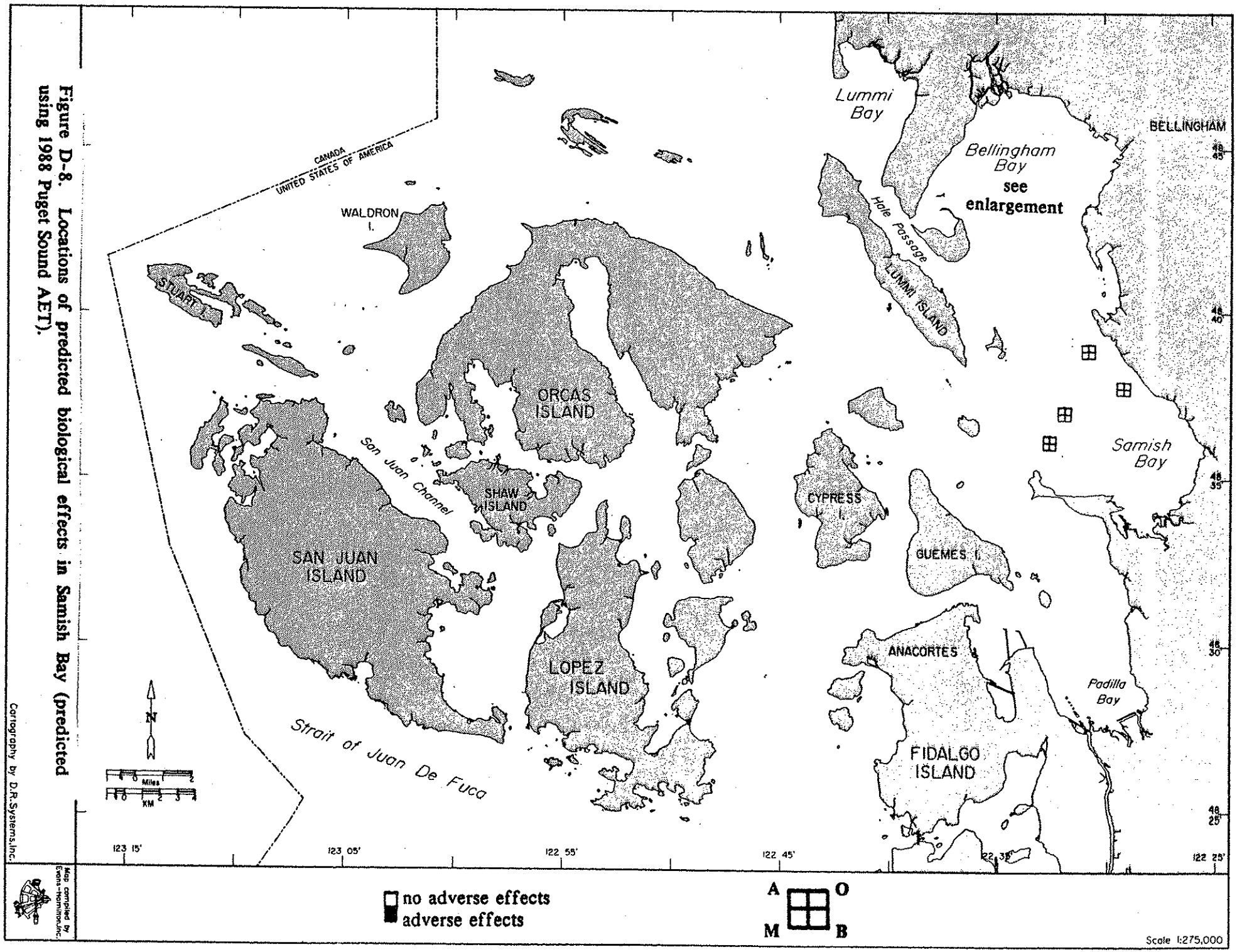


Figure D-8. Locations of predicted biological effects in Samish Bay (predicted using 1988 Puget Sound AET).

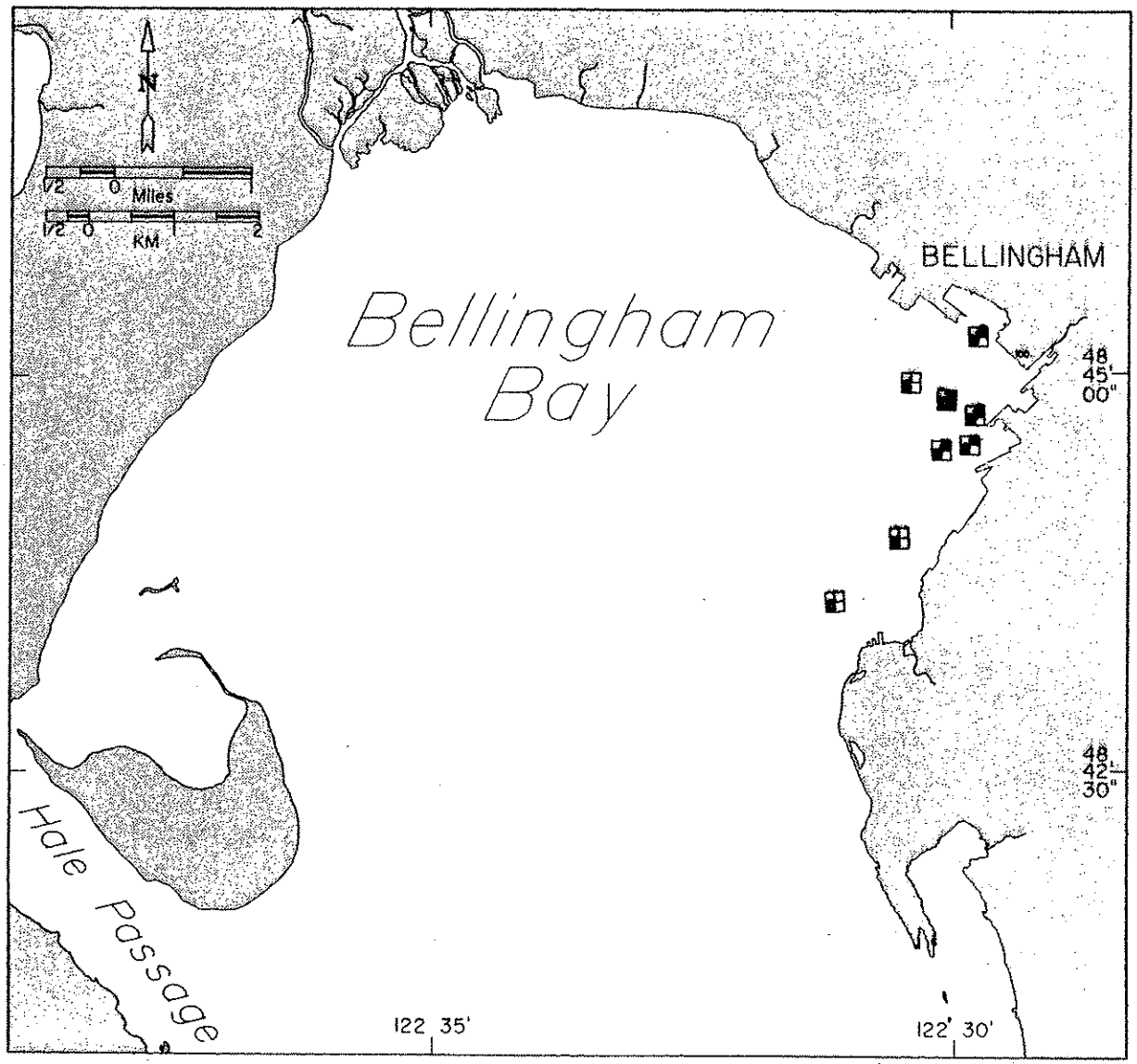


Figure D-9. Locations of predicted biological effects in Bellingham Bay (predicted using 1988 Puget Sound AET).

no adverse effects
adverse effects

A O
M B

Map compiled by Evans-Hamilton, Inc.

Cartography by D.R.Systems, Inc.

Scale 1:84,000

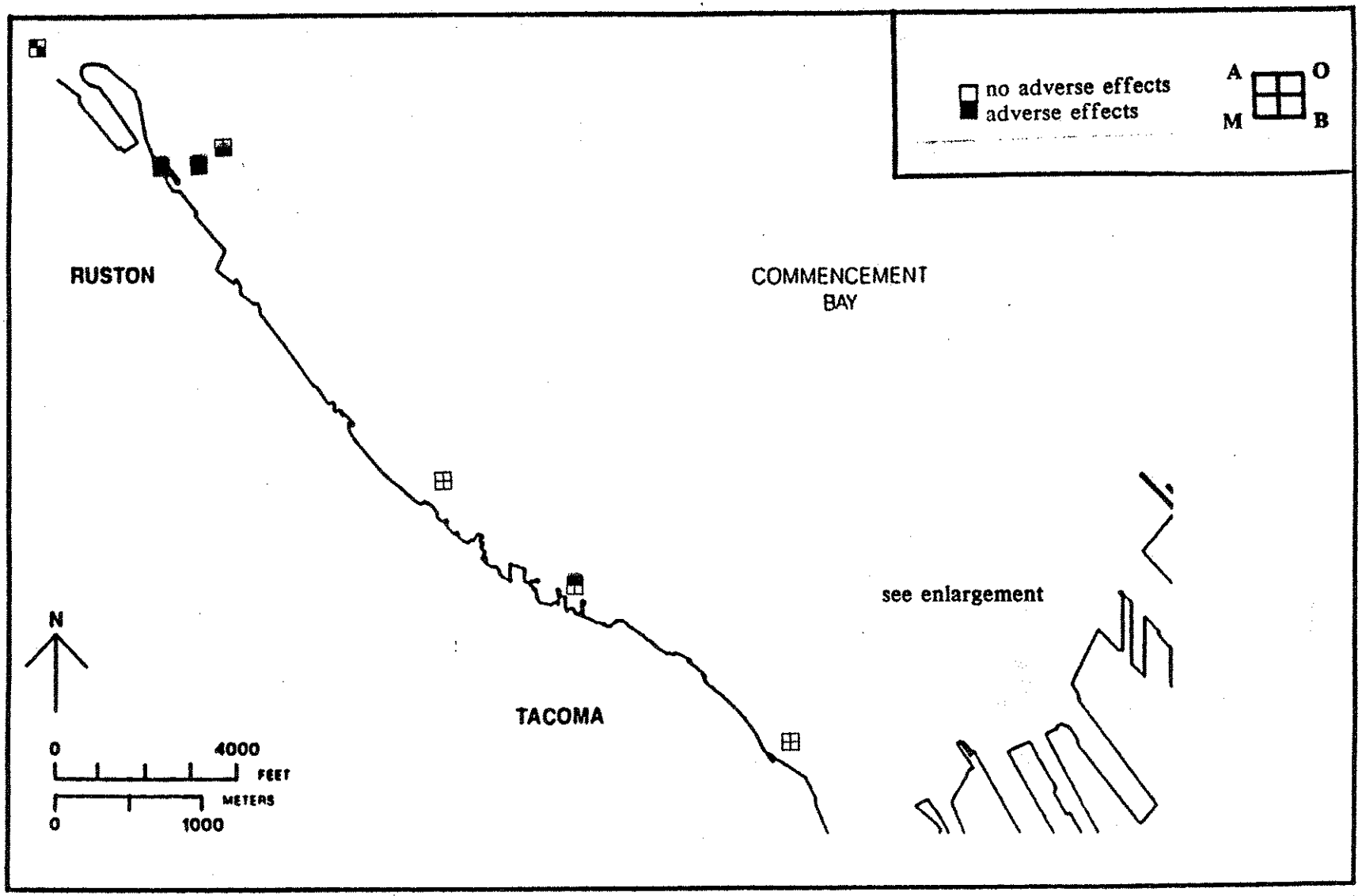


Figure D-10. Locations of predicted biological effects in Commencement Bay - Ruston Shore (predicted using 1988 Puget Sound AET).

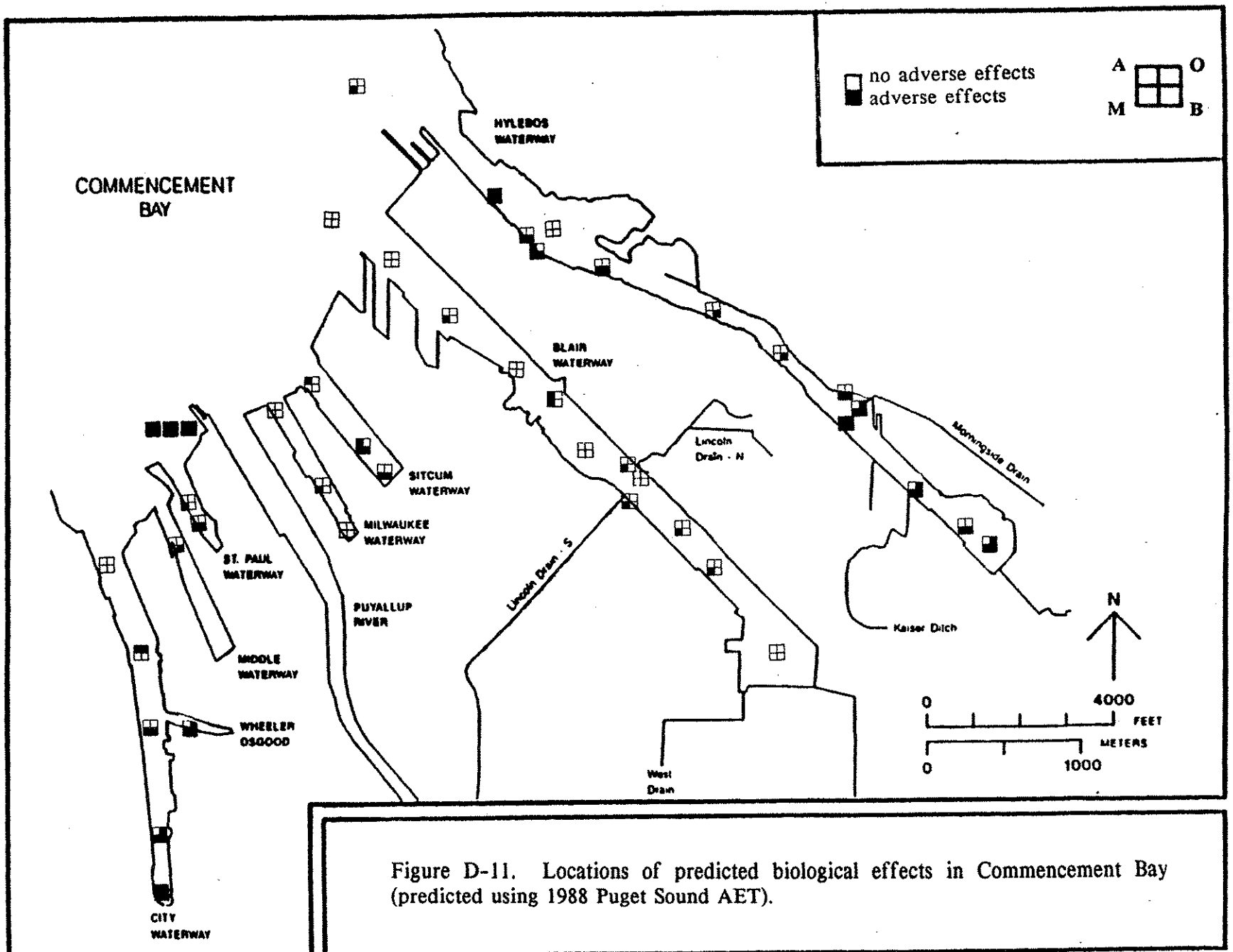


Figure D-11. Locations of predicted biological effects in Commencement Bay (predicted using 1988 Puget Sound AET).

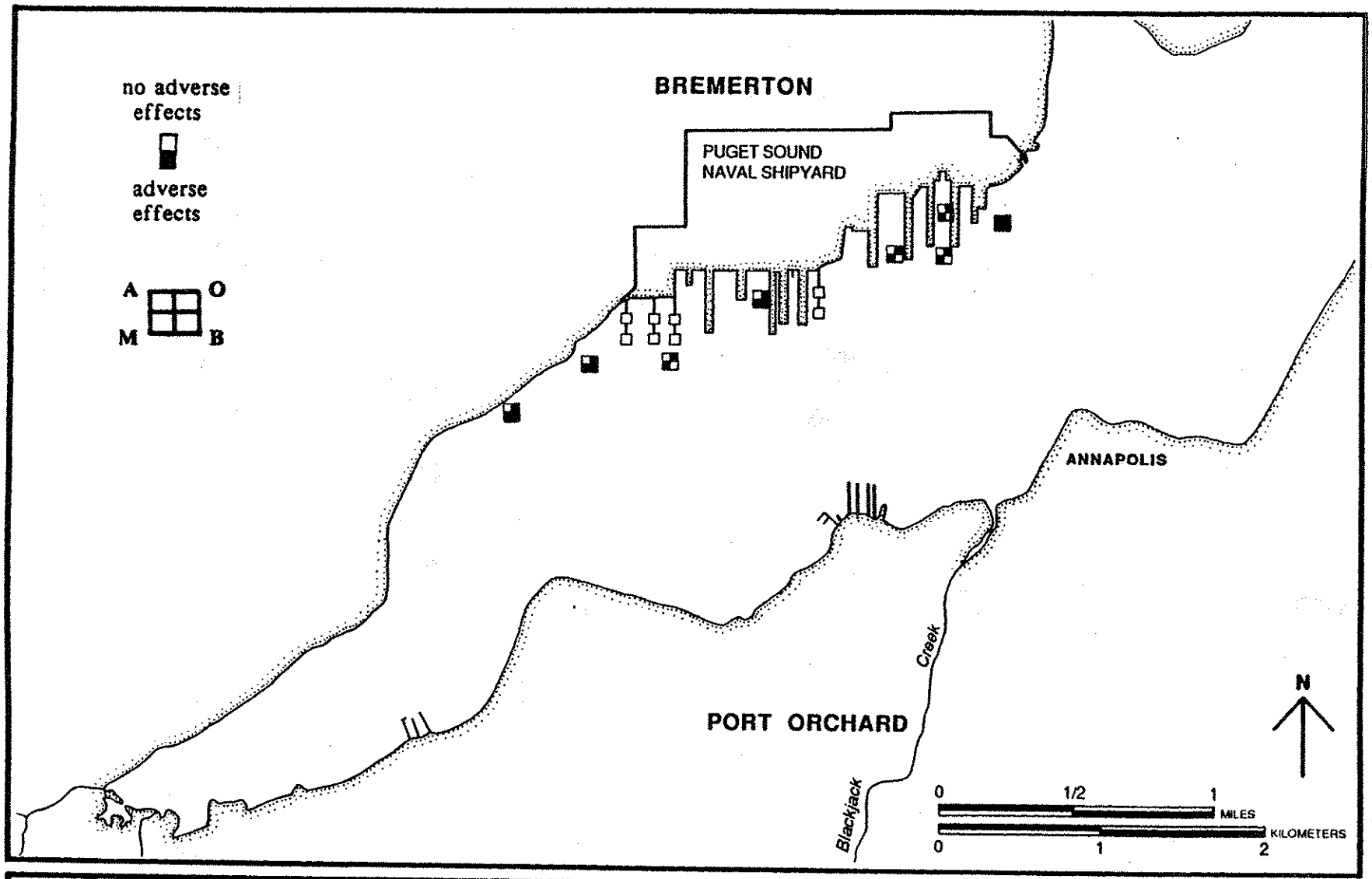


Figure D-12. Locations of predicted biological effects in Sinclair Inlet (predicted using 1988 Puget Sound AET).

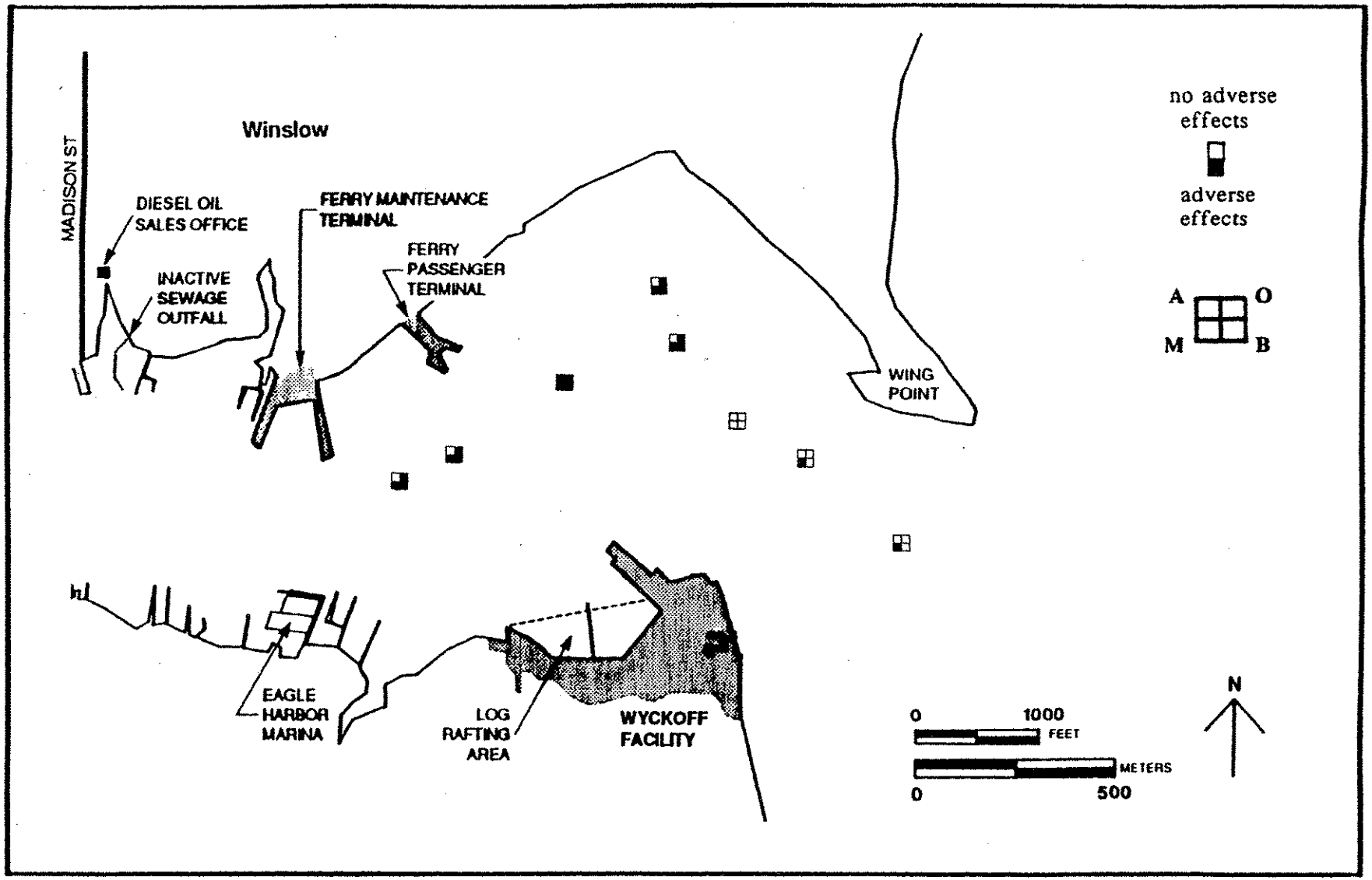


Figure D-13. Locations of predicted biological effects in Eagle Harbor (predicted using 1988 Puget Sound AET).