EPA/620/R-93/006
June 1993

VIRGINIAN PROVINCE DEMONSTRATION REPORT

EMAP-ESTUARIES - 1990

Prepared by

Stephen B. Weisberg
Versar, Inc.
Columbia, MD 21045

A. Frederick Holland
Marine Resources Research Institute
Charleston, SC 29412

K. John Scott
Science Applications
International Corp.
Narragansett, RI 02882

Harold T. Wilson
Coastal Environmental Services, Inc.
Linthicum, MD 21090

Douglas G. Heimbuch
Coastal Environmental Services, Inc.
Linthicum, MD 21090

Steven C. Schimmel
U.S. Environmental Protection Agency
Narragansett, RI 02882

Jeffrey B. Frithsen
Versar, Inc.
Columbia, MD 21045

John F. Paul
U.S. Environmental Protection Agency
Narragansett, RI 02882

J. Kevin Summers
U.S. Environmental Protection Agency
Gulf Breeze, FL 32561

Raymond M. Valente
Science Applications
International Corp.
Narragansett, RI 02882

Jeroen Gerritsen
Versar, Inc.
Columbia, MD 21045

Richard W. Latimer
U.S. Environmental Protection Agency
Narragansett, RI 02882

EPA Contract No. 68-DO-0093

Project Officer

Vernon J. Laurie
Office of Modeling, Monitoring Systems,
and Quality Assurance
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC 20460

NOTICE

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency (Environmental Monitoring and Assessment Program, Office of Research and Development) under contract No. 68-DO-0093 to Versar, Inc. It has been subjected to the Agency's peer and administrative review, and it has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

The suggested citation for this report is: Weisberg, S.B., J.B. Frithsen, A.F. Holland, J.F. Paul, K.J. Scott, J.K. Summers, H.T. Wilson, R. Valente, D.G. Heimbuch, J. Gerritsen, S.C. Schimmel, and R.W. Latimer. 1992. EMAP-Estuaries Virginian Province 1990 Demonstration Project Report. EPA 600/R-92/100. U.S. Environmental Protection Agency, Environmental Research Laboratory, Narragansett, RI.

PREFACE

EMAP is a long-term, integrated, monitoring, research, and assessment program to determine the condition of our nation's ecological resources. The program is organized into types of environmental resources, such as estuaries. An important step in the implementation of EMAP involves conducting regional demonstration projects. This report describes and evaluates an EMAP demonstration project conducted in the estuaries of the mid-Atlantic states in 1990. The project successfully demonstrated that the EMAP sampling design provides statistically valid measures of environmental condition; 1) a suite of biological response, pollutant exposure, and habitat indicators can form the core of future EMAP-Estuaries monitoring activities; 2) the EMAP sampling design provides statistically valid measures of environmental condition; and 3) the assessment techniques appear to be a good way to characterize conditions for resource managers.

Page IV left blank intentionally

EXECUTIVE SUMMARY

The Environmental Monitoring and Assessment Program (EMAP) is a comprehensive environmental monitoring network designed to 1) estimate the current status and trends in the condition of the nation's ecological resources on a regional basis, with known confidence; 2) seek associations between human-induced stress and ecological condition; and 3) provide periodic statistical summaries and interpretive reports on ecological status and trends to resource managers and the public. The program was initiated to provide the information necessary for formulating future environmental policies by answering the following questions: What is the current extent of our ecological resources, and how are they distributed geographically? What proportions of the resources are currently in good or acceptable condition? What proportions are degrading or improving, in what regions, and at what rates? Are these changes correlated with patterns in environmental stresses? Are adversely-affected resources improving overall in response to control and mitigation programs?

Three characteristics of EMAP differentiate it from most previous environmental monitoring programs. First, sampling in EMAP is probability-based so that estimates of status and trends can be made with quantifiable confidence. Second, EMAP monitoring and assessments focus on biological indicators of response to natural and human-induced stress; indicators of pollutant exposure and habitat condition are sampled simultaneously to provide a context for interpreting biological indicators. Third, the scale of EMAP monitoring is regional and national, rather than local.

The program is organized into Resource Groups responsible for conducting assessments of seven categories of environmental resources: forests, wetlands, surface waters, near coastal waters, arid lands, agricultural ecosystems, and the Great Lakes. Administered by the EPA's Office of Research and Development, EMAP is an integrated federal program that is planned and implemented in cooperation with the National Oceanic and Atmospheric Administration, the U.S. Fish and Wildlife Service, the U.S. Forest Service, the U.S. Bureau of Land Management, the U. S. Agricultural Research Service, and the U.S. Geological Survey.

The first stage in implementing EMAP involves conducting demonstration projects for each Resource Group. Demonstration projects provide the opportunity to illustrate the kinds of assessments that can be conducted using EMAP data and to work with users of the data to select the most appropriate indicators for evaluating problems of concern in that resource category. Demonstration projects also provide information for refining the sampling design and identifying and resolving logistical difficulties associated with regional monitoring in each resource category on a limited scale, before EMAP monitoring is implemented nationwide.

This report describes the first EMAP demonstration project, which was conducted in estuaries of the mid-Atlantic states (the Virginian Province) in 1990 by the Near Coastal Resource

Group. Estuaries were selected for the first demonstration project because of their biological productivity and the current intense public interest in restoring and maintaining estuarine resources. The objectives of the project were to demonstrate the utility of regional scale monitoring using a probability-based sampling design for assessing the condition of estuarine resources, establish standardized methods for estuarine monitoring, identify and resolve logistical problems associated with large-scale sampling, test and develop biological indicators of estuarine environmental quality, and collect the data on regional-scale variability in ecological parameters needed to evaluate and refine the sampling design. Each of these objectives is addressed in individual sections of this report. A summary of those sections follows.

The 1990 Demonstration Project involved 500 sampling visits to 217 sites from Cape Cod to the mouth of the Chesapeake Bay. A series of indicators that are representative of the overall health of estuarine resources was measured at each site. These indicators were selected to address three major attributes of concern: 1) biotic integrity, or the existence of healthy, diverse, and sustainable biological communities; 2) pollutant exposure, or the condition of the chemical environment in which biota live; and 3) aesthetics, representing societal values related to public use of estuarine resources. Habitat indicators, such as depth, salinity, temperature, and the physical characteristics of the sediment, were also measured.

METHODS (SECTION 2) AND LOGISTICS (SECTION 3)

One characteristic of the EMAP-Estuaries program is that the entire province is sampled within a limited time period (approximately six weeks) using standardized methods to ensure comparability of data within and among sampling years. One of the goals of the 1990 Demonstration Project was to develop and document these standardized methods and associated quality assurance protocols. This objective was accomplished successfully. Field and laboratory manuals for all EMAP-Estuaries sampling activities are available. EMAP is promoting these methods and associated QA protocols to EPA Regions, states, and local agencies responsible for monitoring to facilitate comparability of information from multiple monitoring programs within the province.

The 1990 Demonstration Project also served as a means for evaluating the logistical feasibility of conducting a regional monitoring program in estuaries and identifying the most difficult obstacles involved therein. The 1990 Demonstration Project was a logistical success. More than 90% of the scheduled samples were collected, and they were all obtained without injury to crew members or significant loss of equipment. More than 95% of the data collected passed QA requirements, and specific alterations to the crew training program were identified to resolve the few data collection problems in future years. Several innovative technologies incorporated into the 1990 Demonstration Project for evaluation were deemed a success, such

as the use of on-board computers in small boats to record station position at the time of sampling, and the use of bar-coding to track sample shipments to the numerous laboratories responsible for processing.

INDICATOR DEVELOPMENT AND TESTING (SECTION 4)

A major accomplishment of the 1990 Demonstration Project was developing and applying a methodology for calibrating and validating biological indicators of estuarine condition. The methodology uses discriminant analysis to identify the most effective combination of measurements for distinguishing assemblages at regional reference sites from those at sites with known environmental perturbations. Such biological indicators, which are applicable over a range of latitudes and habitats, had not been identified previously. Developing these indicators is important for EMAP and should prove helpful in other Agency efforts for developing biocriteria.

Benthic invertebrate assemblages were the most successful biological indicators. Bottom-dwelling organisms are particularly useful as indicators because they integrate exposure conditions over long periods of time (months to years), and because their relative immobility prevents them from avoiding pollution exposure. Based on studies conducted during the 1990 Demonstration Project, five attributes of the assemblage related to the number of species, number of amphipods, number of capitelid polychaetes, number of bivalves, and the average weight per polychaete can correctly differentiate clean reference sites from polluted sites with about 90% certainty. Initial steps have been taken to validate this index; however, additional data from future years of the program will be used to further refine the index and continue the validation process.

Fish assemblage indicators also are promising for distinguishing between polluted and unpolluted sites, but additional data will be required to reduce uncertainty. The mobility of fish presents the biggest obstacle to using them for identifying conditions at a particular site. Even if a fish indicator that successfully discriminates sites of high and low quality cannot be validated until later years of the program, there are still several fish parameters, such as prevalence of visible pathological disorders, that convey meaningful information on a regional or watershed basis.

Measures of pollution exposure also were evaluated during the 1990 Demonstration Project. For instance, variability in dissolved oxygen concentrations in the province was examined by deploying continuously-recording meters for periods of up to 60 days at a subset of sites. Although DO was found to be quite variable at individual sites, it was stable enough on a regional basis for making provincewide estimates of condition. Improved methods for measuring dissolved oxygen on a regional scale are being developed and tested in 1991 based on the findings of the 1990 Demonstration Project.

DESIGN EVALUATION (SECTION 5)

One of the primary objectives of the 1990 Demonstration Project was to determine the precision with which status can be estimated using the present design. At the province level, confidence intervals surrounding status estimates for each indicator measured during the 1990 Demonstration Project were less than 10% of the areal extent of the province. The uncertainty associated with these estimates is anticipated to decrease substantially after completing a four-year sampling cycle. Estimates developed at the subprovince level, such as for individual estuarine systems like the Chesapeake Bay, were less precise because fewer samples were collected in the subpopulations, but estimates were still within 15% of the actual areal extent. Confidence intervals for these subpopulations will also contract when the four-year sampling cycle is completed.

Another objective of the Demonstration Project was to estimate sources of variability for a selected indicator (the benthic index) and incorporate this information into a power analysis model to assess its power for trend detection using the present sampling design and sample density. Estimates of variability were generated from two sources: the EMAP sampling program and an ongoing monitoring program in the lower Chesapeake Bay. The power for trend detection was found to be sensitive to the estimate for interannual variability. Using the estimate generated from EMAP data, the likelihood of detecting a 2% per year change in the area of degraded benthic invertebrate assemblages over 12 years was greater than 99%. Using the lower Chesapeake Bay data set to estimate variance, the power for detecting a 2% per year change was only about 60%. Neither data set was ideal for estimating interannual variability; the EMAP data set was limited temporally, and the Chesapeake Bay data set was limited spatially. As EMAP continues to collect data over the next several years, the power analysis will be refined to better define the power of the program for detecting trends over decades.

The stability of several indicators was evaluated across three sampling intervals (mid-June to late-July; late-July to the end of August; early to late September) to define the boundaries of the sampling window. The most appropriate sampling period was determined to be late July through August because 1) dissolved oxygen values are at annual low values; 2) contaminant exposure is greatest because of low dilution flows and peak metabolic activity associated with highest water temperatures; and 3) living resources are most abundant, maximizing the probability of collecting organisms required for assessments.

Sample allocation in the 1990 Demonstration Project was accomplished after stratifying estuaries into three classes (large estuaries, large tidal rivers, and small estuarine systems). Without this stratification, the large tidal river and small estuarine system classes, which are perceived to be at risk from different types of stresses, would not have been sampled sufficiently to make assessments with acceptable levels of uncertainty. Alternative

stratification schemes based on more dynamic characteristics, such as salinity and sediment type, were examined and found to have logistical shortcomings that rendered them less appropriate than the present design.

Another design alternative examined was the use of index sites in large tidal rivers and in small estuarine systems. Whereas random sites are designed to estimate areal extent, index sites are located in the areas of a system most likely to exhibit a problem if one exists (deep, depositional sites) and are used to estimate the percent of systems experiencing environmental degradation. Index sites added little value to the program. The biggest impediment to using index sites was that existing sediment maps of the Virginian Province are inadequate to define depositional areas accurately.

PRELIMINARY EVALUATION OF ESTUARINE STATUS (SECTION 6)

When fully implemented, EMAP will provide regional and national assessments of ecological status and trends for the nation's environmental resources based on a four-year sampling cycle. The multi-year baseline is intended to minimize the effect of natural interannual variability due to climate and other factors. The preliminary evaluation of the condition of estuaries in the mid-Atlantic region provided in this report represents a first attempt at presenting statistically unbiased, regional-scale information to a broad audience and is intended to elicit discussion about assessment needs. The evaluation is preliminary because some indicators have not been validated fully, a process that will require several years of regional-scale data. The estimates presented are based on a single year of data rather than the four-year running average that is the basic unit of EMAP assessment; nonetheless, province-level evaluations of ecological condition that are unavailable from other sources are possible with the data:

- The biotic integrity of estuaries in the Virginian Province was evaluated by measuring the condition of benthic invertebrate (bottom-dwelling animals) assemblages. Between 16%-30% of the estuarine area in the Province had benthic resources that were degraded compared to regional reference sites.

- Biotic integrity was also assessed by examining the prevalence of visible pathological disorders (lesions, tumors, etc.) of fish. Four of every thousand fish in the province had a visible pathological disorder. The prevalence in demersal fish (those living in close contact with the bottom sediments) was several times that in pelagic fish (those living primarily in the water column). Less than 0.1% of fish that are commercially or recreationally harvested had visible pathological disorders.

- Ten-day solid-phase toxicity tests using indigenous biota were conducted to examine the condition of estuarine sediments. Eight percent of the sediments in the province were estimated to be acutely toxic.

- Sediments were screened for contaminants using the same list of analytes used in the NOAA National Status and Trends Program. Based on this list of analytes, 39% of the province was estimated to have concentrations of contaminants in the sediments that potentially cause at least sublethal biological effects. Metals, lead and zinc in particular, were the most prevalent contaminants at these concentrations.

- High concentrations of *Clostridium perfringens*, a bacterial tracer of sewage pollution, were found in an estimated 9% of the province.

- Small estuarine systems, including harbors, bays and coastal embayments, had the highest proportion of toxic sediments, sediments containing contaminant concentrations of biological concern, and sediments containing high levels of *Clostridium*. They also had the highest proportion of fish with pathological disorders. These small systems typically are overlooked in monitoring programs that concentrate effort along the main axis of large estuarine systems.

- Between 14%-28% of the area in the province had dissolved oxygen concentrations below 5 ppm, the water quality standard for many states in the province. Nine percent of the area was estimated to have concentrations below 2 ppm, which is considered stressful to most biota.

- Of the largest systems in the Virginian Province, Long Island Sound had the highest proportion of area with oxygen concentrations less than 5 ppm; Chesapeake Bay had the highest proportion of area with concentrations below 2 ppm.

- Anthropogenic marine debris (trash) was estimated to be present in 9-19% of the estuarine area of the province. Paper and plastic accounted for most of this debris.

CONTENTS

Page

NOTICE ii

PREFACE iii

EXECUTIVE SUMMARY v

CONTENTS xi

FIGURES xv

TABLES xix

ABBREVIATIONS xxii

1. INTRODUCTION 1-1

1.1 OVERVIEW OF EMAP 1-1
1.2 INTRODUCTION TO 1990 DEMONSTRATION PROJECT 1-2
1.3 OBJECTIVES OF THE 1990 DEMONSTRATION PROJECT 1-4
1.4 PURPOSE AND ORGANIZATION OF THIS REPORT 1-5

2. DEMONSTRATION PROJECT APPROACH AND METHODS 2-1

2.1 SAMPLING DESIGN 2-3

2.1.1 Base Sampling Sites 2-4
2.1.2 Long-term Dissolved Oxygen (LTDO) Sites 2-5
2.1.3 Supplemental Sampling Sites 2-5
2.1.4 Indicator Testing and Evaluation Sites 2-6
2.1.5 Index Sites 2-6

2.2 SAMPLING METHODS 2-6

2.2.1 Bottom Dwelling (Benthic) Animals 2-8
2.2.2 Sediment Profile Images 2-14
2.2.3 Fish 2-15
2.2.4 Large Bivalve Abundance and Tissue Contamination 2-18

Table of Contents (*Cont.*)

2.2.5 Dissolved Oxygen Concentration 2-19
2.2.6 Chlorophyll *a* 2-20
2.2.7 Sediment Toxicity 2-20
2.2.8 Sediment Contaminant Concentrations 2-21
2.2.9 Sediment *Clostridium* Spores 2-21
2.2.10 Sediment Grain Size 2-23
2.2.11 Water Column Toxicity 2-24
2.2.12 Marine Debris 2-24

2.3 DATA COLLECTION AND SAMPLE TRACKING 2-25

2.4 ANALYTICAL METHODS FOR STATUS ASSESSMENT 2-25

2.4.1 Assessment Methods for Individual Indicators 2-26
2.4.1.1 Benthic Index 2-26
2.4.1.2 Fish 2-28
2.4.1.3 Sediment Toxicity 2-28
2.4.1.4 Sediment Contaminant Concentrations 2-29
2.4.1.5 Sediment *Clostridium* Spores 2-29
2.4.1.6 Marine Debris 2-30
2.4.1.7 Water Clarity 2-30
2.4.1.8 Water Column Stratification 2-30
2.4.1.9 Integration of Indicators 2-30

2.4.2 Assessment Methods for Areal Estimation 2-31
2.4.3 Procedures for Estimating Precision 2-33

3. EVALUATION OF LOGISTICAL FEASIBILITY AND QUALITY ASSURANCE 3-1

3.1 WATER QUALITY PARAMETERS 3-3
3.2 SEDIMENT QUALITY AND BENTHIC COMMUNITY PARAMETERS 3-5
3.3 FISH AND BIVALVE SAMPLING 3-12

4. INDICATOR DEVELOPMENT AND TESTING 4-1

4.1 RESPONSE INDICATORS 4-2

4.1.1 General Methodology for Developing Response Indicators 4-4

Table of Contents (*Cont.*)

4.1.2 Benthic Macroinvertebrates 4-10
4.1.2.1 Development of a Benthic Index of Estuarine Integrity 4-10
4.1.2.2 Discussion of the Benthic Index 4-22
4.1.2.3 Sediment Profile Imaging Techniques 4-23
4.1.3 Fish 4-27
4.1.3.1 Development of a Fish Index of Estuarine Quality 4-26
4.1.3.2 Discussion of the Fish Index 4-31
4.1.3.3 Visible Pathology 4-33
4.1.3.4 Fish Tissue Contaminants 4-34

4.2 EXPOSURE INDICATORS 4-36

4.2.1 Dissolved Oxygen 4-36

4.2.1.1 Use of dissolved oxygen data for estimating status and trends 4-38
4.2.1.2 Use of the dissolved oxygen indicators for associations 4-43

4.2.2 Sediment Toxicity 4-51
4.2.3 Sediment Contaminants 4-54
4.2.3.1 Identifying Anthropogenic Enrichment 4-54
4.2.3.2 Identifying "Above Background" Concentrations 4-58
4.2.3.3 Determining Biologically Significant Concentrations 4-59
4.2.3.4 Comparison of Analytical Approaches 4-60

4.3 HABITAT INDICATORS 4-68

4.3.1 Depth 4-68
4.3.2 Salinity 4-68

5. EVALUATION OF THE SAMPLING DESIGN 5-1

5.1 DETERMINATION OF THE APPROPRIATE TIME FRAME FOR SAMPLING . 5-1
5.2 STRATIFIED SAMPLING 5-3
5.3 PRECISION OF STATUS ESTIMATES 5-12

5.3.1 Improving the Precision of Estimates 5-17
5.3.2 Using Replicates to Improve Precision 5-18
5.3.3 Improving Estimation of Precision 5-21

Table of Contents ***(Cont.)***

5.4 EVALUATION OF POWER FOR TREND DETECTION 5-22

5.4.1 Formulation of the Model 5-24
5.4.2 The Trend Detection Procedure 5-25
5.4.3 Component Parameter Estimates 5-28
5.4.4 Results from the Trend Detection Model 5-30

5.5 INDEX SITES 5-32
5.6 DETERMINATION OF THE APPROPRIATE SPATIAL SAMPLING SCALE . 5-34

6. STATUS OF VIRGINIAN PROVINCE ESTUARIES 6-1

6.1 INTRODUCTION 6-1
6.2 BIOTIC INTEGRITY 6-3
6.3 POLLUTANT EXPOSURE 6-7
6.4 SPECIFIC ESTUARINE SYSTEMS 6-12
6.5 SPECIALIZED HABITATS 6-19
6.6 ASSOCIATIONS 6-21
6.7 SOCIETAL VALUES 6-22
6.8 INTEGRATION OF ESTUARINE CONDITIONS 6-25
6.9 PROGRAM DIRECTION 6-26

7. SUMMARY AND CONCLUSIONS 7-1

8. REFERENCES 8-1

FIGURES

Figure 2-1 EMAP biogeographical provinces for estuaries 2-2

Figure 2-2 Schematic summarizing the indicator testing and evaluation strategy for the 1990 Demonstration Project in the Virginian Province 2-7

Figure 2-3 Areas sampled by each team during the 1990 Demonstration Project in the Virginian Province 2-9

Figure 2-4 Overview of the EMAP indicator strategy giving examples of the types of indicators used to assess estuarine status 2-10

Figure 4-1 General methodology used to evaluate potential response indicators 4-5

Figure 4-2 Plot showing the regression line used to estimate salinity-adjusted species richness measures for mean number of species per site 4-14

Figure 4-3 Relationship of number of species to salinity after adjusting for habitat variables 4-16

Figure 4-4 Time series data obtained at two long-term dissolved oxygen monitoring sites showing fluctuation of dissolved oxygen concentrations over the summer index period 4-38

Figure 4-5 Comparison of CDFs for province-wide botom dissolved oxygen concentrations in sampling intervals 2 and 3 4-41

Figure 4-6 Example showing the degree of autocorrelation on samples from dissolved oxygen concentration 4-42

Figure 4-7 Differences in dissolved oxygen due to diurnal and tidal periodicities at stations of different depths 4-43

Figure 4-8 CDFs developed by restricting sampling of time series data to daylight hours and by unrestricted random sampling 4-45

Figure 4-9 Relative performance of the six sampling strategies in estimating the mean dissolved oxygen concentration within 1 ppm from time series 4-47

Figure 4-10 Dissolved oxygen quantities from minimum to 25%, for 13 LTDO data series 4-48

Figure 4-11 Misclassification rate associated with DO threshold values for one sampling strategy 4-49

Figure 4-12 Overall misclassification rates for the point-in-time sampling (a) and short-term continuous sampling (b) derived from plots like the one shown in Figure 5-13 4-50

Figure 4-13. Relationship between chromium and aluminum from 1990 EMAP data . . . 4-57

Figure 4-14. Percent of area in each of three resource classes with elevated contaminant concentrations, as defined by four different threshold approaches 4-68

Figure 4-15 Comparison of cumulative distribution function for depth from sampling intervals 2 and 3 4-70

Figure 4-16 Comparison of cumulative distribution function for salinity from sampling intervals 2 and 3 4-71

Figure 5-1 Comparison of cumulative distribution function for dissolved oxygen between sampling intervals 2 and 3 5-6

Figure 5-2 Hypothetical distribution functions used in analysis of stratum misclassification effects 5-10

Figure 5-3 Comparison of optimal vs proportional allocation 5-11

Figure 5-4 Confidence intervals for estimates of benthic index and sediment toxicity over the entire Virginian Province 5-13

Figure 5-5 Confidence intervals for estimates of the benthic index and sediment toxicity by resource classes 5-14

Figure 5-6 Confidence intervals for estimates of the benthic index and sediment toxicity by resource system 5-15

Figure 5-7 Comparisons of CDFs for the benthic index between index and random sites in small estuaries and tidal rivers 5-33

Figure 5-8 Comparison of CDFs for sediment type, benthic index, depth, and sediment toxicity, between the base and intensified sampling efforts in the Delaware Estuary 5-35

Figure 6-1. Location of the EMAP 1990 Demonstration Project sample sites in the estuaries of the Virginian Province 6-2

Figure 6-2. Percent of estuarine area in the three classes of estuarine resources having degraded benthic communities 6-5

Figure 6-3. Prevalence of fish with pathological disorders in the Virginian Province 6-7

Figure 6-4. Prevalance of pathological disorders in demersal fish in the three classes of estuarine resources 6-8

Figure 6-5. Percent of estuarine area in the Virginian Province with bottom dissolved oxygen concentrations below 2 ppm and 5 ppm 6-9

Figure 6-6. Percent of area in the Virginian Province with sediment contaminant concentrations greater than Long and Morgan ER-L values 6-11

Figure 6-7. Pollutant exposure conditions for the three classes of estuarine resources 6-13

Figure 6-8. Sediment types in the three major estuarine systems in the Virginian Province 6-14

Figure 6-9. Salinity habitats in the three major estuarine systems in the Virginian Province: brackish (0 to 5 ppt), transitional (5 to 18 ppt), and marine (> 18 ppt) 6-15

Figure 6-10. Bottom dissolved oxygen conditions in the three major estuarine systems in the Virginian Province 6-16

Figure 6-11. Prevalence of four classes of contaminants in the three largest estuarine systems in the Virginian Province 6-17

Figure 6-12. Degraded benthic resources in the three largest estuarine systems in the Virginian Province 6-18

Figure 6-13. Salinity habitats in the Virginian Province as a percent of total estuarine area 6-20

Figure 6-14. Association between degraded benthic assemblages and the dissolved oxygen and sediment toxicity pollutant exposure indicators 6-22

Figure 6-15. Percent of area having anthropogenic marine debris (trash) in three classes of estuarine resources 6-24

Figure 6-16. Percent of area in the three classes of estuarine resources where high concentrations of *Clostridum perfringens* were found in the sediment 6-25

Figure 6-17. Conceptual scheme for combining indicators that might be measured by EMAP into an integrated statement about the environmental condition of estuaries 6-27

Figure 6-18. Summary of environmental conditions in the estuaries of the Virginian Province 6-28

TABLES

2-1 Summary of the characteristics of estuarine classes in the Virginian Province 2-5

2-2 Sampling activities accomplished at each type of station in each of the three sampling intervals during the 1990 Demonstration Project 2-11

2-3 Biomass groups for macrofaunal species 2-13

2-4 Calculation of the OSI based on sediment profile images 2-15

2-5 Targeted species and size ranges of fish retained for tissue analysis during the 1990 Virginian Province Demonstration Project 2-16

2-6 Analytical measurements for fish and bivalve tissue samples collected during the 1990 Virginian Province Demonstration Project 2-17

2-7 Analytical measurements for sediment samples collected during the 1990 Virginian Province Demonstration Project 2-22

2-8 Analytical methods used in 1990 for determination of chemical contaminant concentrations in sediments 2-23

2-9 Standardized mean and standard deviation for indicators used in the discriminant function for the benthic index 2-27

3-1 Status of sample collection during the 1990 Virginian Province Demonstration Project 3-2

3-2 Results for SRM 2704 used as a set control for the 1990 EMAP-E sediment inorganic analyses 3-7

3-3 Results for SRM 1941 used as a set control for the 1990 EMAP-E sediment organic analyses 3-8

3-4 Range in detection limits reported for organic compounds in 1990 sediment samples 3-9

4-1 Test for significant differences between degraded and nondegraded reference sites for each of the descriptors of the benthic assemblage 4-13

4-2 Summary of correlation between habitat indicators and the candidate benthic measures 4-15

4-3 List of sites used in indicator testing dataset during the 1990 Virginina Province Demonstration Project 4-18

4-4 Number of indicator testing sites sampled in each of several habitat types 4-19

4-5 Results of discriminant analyses conducted to combine candidate benthic measures into an index 4-20

4-6 Results of validation of four candidate indices 4-22

4-7 Correlation coefficients relating the values of each of four indices and percent of sites that were classified the same among four indices applies to 152 sampling sites 4-23

4-8 Comparison of benthic characterizations produced by sediment profile image analysis and conventional macroinvertebrate sampling 4-26

4-9 Tests for significant differences among degraded and reference sites for each candidate fish measure 4-28

4-10 Results of discriminant analyses conducted for combining candidate fish measures into an index 4-31

4-11 Correlations between the first and second trawl and between intervals 2 and 3 for each candidate fish measure 4-33

4-12 Relationship of contaminant concentration to Long and Morgan ER-M values and toxicity by site, as a function of salinity, grain size, and holding time 4-53

4-13 Relationship between sediment metal concentration and aluminum 4-58

4-14 Comparison of threshold contaminant levels based upon the "above background", ER-L, and ER-M values 4-62

4-15 Percent of the area in the Virginian Province exceeding each of four threshold sediment contaminant concentrations 4-65

5-1 T-test comparison of dissolved oxygen concentrations and the benthic index among periods 5-3

5-2 Average dissolved oxygen concentration by week for stations with relatively complete Hydrolab records 5-4

5-3 Weekly minimum dissolved oxygen concentrations for stations with relatively complete Hydrolab records 5-5

5-4 Comparison of sediment type predicted from NOAA database with that measured for EMAP sampling sites 5-8

5-5 Confidence intervals associated with estimates around critical values of response and exposure indicators 5-16

5-6 Optimal percent of sample size devoted to annual sampling 5-23

5-7 Estimates of variance components used in the power anlaysis based on EMAP data 5-29

5-8 Power for trend detection in the benthic index over three EMAP sampling cycles 5-31

ABBREVIATIONS

ACE	U.S. Army Corps of Engineers
ANOVA	analysis of variance
APHA	American Public Health Association
ARS	Agricultural Research Service
BI	benthic index
BLM	Bureau of Land Management
CDF	cumulative distribution function
CEQ	Council on Environmental Quality
CPR	cardio-pulmonary resuscitation
CTD	conductivity, temperature, depth
DO	dissolved oxygen
EMAP	Environmental Monitoring and Assessment Program
EMAP-E	EMAP Estuaries Resource Group
EPA or USEPA	U.S. Environmental Protection Agency
ER-L	effects range-low value
ER-M	effects range-median value
FDA	Food and Drug Administration
FS	U.S. Forest Service
FWS	U.S. Fish and Wildlife Service
GFAA	graphite furnace atomic adsorption
IBI	Index of Biotic Integrity

ICP-AES	inductively coupled plasma-atomic emission spectrometry
ITE	Indicator Testing and Evaluation
LTDO	Long-term Dissolved Oxygen
loran	long range navigation
MSE	mean squared error
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
ORD	Office of Research and Development
OSI	Organism-Sediment Index
OTA	Office of Technology Assessment
PAH	polycyclic aromatic hydrocarbons
PAR	photosynthetically active radiation
PCB	polychlorinated biphenyls
ppm	parts per million
ppt	parts per thousand
QA/QC	quality assurance/quality control
REMOTS®	Remote Ecological Monitoring of the Seafloor
RPD	redox potential discontinuity
SDD	Secchi disc depth
SRM	standard reference material
USGS	U.S. Geological Survey

SECTION 1
INTRODUCTION

Pollution control programs in the United States have been estimated to cost more than $80 billion annually (CEQ 1990). Most of these programs address specific local pollution problems; however, the means to assess the effectiveness of these programs for protecting the environment at national and regional scales and over extended periods of time do not exist. The U.S. Environmental Protection Agency (EPA) considers it critical to establish monitoring, research, and assessment programs to determine the effectiveness of pollution control strategies and to substantiate the science upon which these strategies are based (USEPA 1992).

1.1 OVERVIEW OF EMAP

The Environmental Monitoring and Assessment Program (EMAP) is a nationwide program initiated by EPA's Office of Research and Development (ORD). EMAP was developed in response to the demand for information about the degree to which existing pollution control programs and policies protect the nation's ecological resources. EMAP is an integrated federal program; ORD is coordinating, planning, and implementing EMAP in conjunction with other federal agencies, including the Agricultural Research Service (ARS), the Bureau of Land Management (BLM), the U.S. Fish and Wildlife Service (FWS), the U.S. Forest Service (FS), the U.S. Geological Survey (USGS), and the National Oceanic and Atmospheric Administration (NOAA), and with other offices within EPA (e.g., Office of Water). These other agencies and offices participate in the collection and analysis of EMAP data and will use it to guide their policy decisions, as appropriate.

The objectives of EMAP are to estimate the current status and trends in the condition of the nation's ecological resources on a regional basis, with known confidence; to seek associations between human-induced stresses and ecological condition; and to provide periodic statistical summaries and interpretive reports on ecological status and trends to resource managers and the public. It is designed to provide the information required to formulate the environmental protection policies of the 1990s and beyond by providing answers to the following questions:

- What is the status, extent, and geographical distribution of the nation's ecological resources?
- What proportion of these resources is declining or improving? Where? At what rate?

- What factors are likely to be contributing to declining condition?
- Are pollution control, reduction, mitigation, and prevention programs achieving overall improvement in ecological condition?

Assessment of status and trends in the condition of the nation's ecological resources requires collecting data in a standardized manner, over large geographic scales, and for long periods of time. Such assessments cannot be accomplished by aggregating data from the many individual, short-term monitoring programs that have been conducted in the past or are being conducted currently (NRC 1990). Differences in the parameters measured, collection methods, timing of sample collection, and program objectives severely limit the value of historical monitoring data and existing monitoring programs for making integrated regional and national assessments.

EMAP was proposed and developed by EPA/ORD because an integrated monitoring and assessment program that samples ecological resources in a probability-based manner offers considerable advantages over historical monitoring approaches. One advantage is improved definition of the extent and magnitude of pollution problems at regional and national scales. Simultaneous, probability-based sampling of pollution exposure, environmental condition, and biological resources, however, is most important. This approach enables estimates to be made of the uncertainty associated with assessments and will improve our ability to identify ecological responses to pollution. EMAP, therefore, will provide objective assessments of the severity and extent of environmental problems and the degree to which degraded resources are responding to efforts to protect or restore them.

1.2 INTRODUCTION TO 1990 DEMONSTRATION PROJECT

EMAP is being initiated in phases, beginning in 1990 with demonstration and pilot projects in the estuaries of the Virginian Province (i.e., the mid-Atlantic region) and the forests of New England. These projects afford EMAP the opportunity to refine the selection of indicators and evaluate and modify the sampling design. They also provide the opportunity to resolve difficulties associated with regional-scale sampling. Demonstration projects combine the elements of several pilot projects (such as testing selected indicators and specific design aspects) with a demonstration of the EMAP approach to monitoring and assessment of ecological condition.

Estuaries were selected as one of the first resources to be sampled by EMAP. Estuaries are among the most productive of ecological systems. Historically, more than 70% of commercial and recreational landings of fish and shellfish were taken from estuaries (NOAA 1987). Estuaries also provide feeding, spawning, and nursery habitats, and migratory routes for many commercially and recreationally important fish and wildlife, including threatened and

endangered species (Lippson et al. 1979; Olsen et al. 1980). The public values estuarine ecosystems for recreation (e.g., boating, swimming, hunting, and fishing) and aesthetic appeal. Approximately $7 billion in public funds are spent annually on outdoor marine and estuarine recreation in the 33 coastal states (NOAA 1988). Millions of tourists visit coastal beaches annually, and coastal property is among the nation's most valuable real estate. About half of the nation's population now lives in coastal areas, and by the year 2010, the population in these areas will have grown by almost 60% to 127 million people (Culliton et al. 1990). The estuarine and coastal environment also provides cooling waters for energy production, transportation routes for ships, and space for economic development. Most of the nation's major ports are located in estuaries.

Estuaries are complex transition zones between streams, rivers, and coastal oceans. They have physical features that concentrate and retain pollutants, and they tend to serve as repositories for the many pollutants released into the nation's waters and into the atmosphere. The ecological condition of estuaries is.influenced strongly by human activities in the watershed, particularly land use patterns and the release of pollutants to the environment. In many coastal regions, water and sediment quality and the abundance of living resources are perceived to have declined over the past 10 to 15 years, despite the implementation of more strict pollution control programs. Increasingly, reports appear in the popular press and scientific journals (Morganthau 1988; Toufexis 1988; Smart et al. 1987) describing the decline of estuarine and coastal environmental quality, as exemplified by:

- Increases in the frequency, duration, and extent of water containing insufficient oxygen to sustain living resources (USEPA 1984; Officer et al. 1984; Parker et al. 1986; Rabalais et al. 1985; Whitledge 1985);

- Accumulation of contaminants in bottom sediments and in the tissues of fish and shellfish to levels that threaten human health and the sustainability of fish and shellfish populations (OTA 1987; NRC 1989);

- Increased evidence that many restoration and mitigation efforts have not replaced losses of critical habitats (Sanders 1989; The Conservation Foundation 1988);

- Increased incidence of pathological problems in fish and shellfish (Sinderman 1979; O'Connor et al. 1987; Buhler and Williams 1988; Capuzzo et al. 1988);

- Increased frequency and persistence of algal blooms and associated decreases in water clarity (USEPA 1984; Pearl 1988; Smayda and Villareal 1989);

- Increased incidence of closures of beaches, shellfishing grounds, and fisheries because of pathogenic and chemical contamination (Smart et al. 1987; FDA

1971, 1985; Hargis and Haven 1988; Broutman and Leonard 1988; Leonard et al. 1989); and

- Increased incidence of human health problems from consuming contaminated fish and shellfish or swimming in contaminated waters (Fein et al. 1984; Malins 1989).

The Virginian Province was selected as the testing ground for the EMAP monitoring effort for estuaries because there is a general public perception that estuaries in this region of the country are deteriorating more rapidly than in other regions. Many of the estuaries in this province have been investigated intensively by scientists, and a considerable amount of information was available for use in designing the 1990 Demonstration Project. Six National Estuary Programs are in place in the Virginian Province. In addition, many management decisions are forthcoming, including development of a restoration plan for the New York Harbor complex, and development of management plans and evaluation of previous management actions for many large estuaries, including Delaware Bay, Chesapeake Bay, and Long Island Sound. Development of such plans presents an opportunity to demonstrate how EMAP monitoring data can assist in the formulation of environmental programs and policies.

1.3 OBJECTIVES OF THE 1990 DEMONSTRATION PROJECT

The specifics of the 1990 Virginian Province Demonstration Project are documented in Holland (1990). A critical issue addressed during the 1990 Demonstration Project was how best to represent the ecological condition of estuarine resources on a regional scale with the limited financial resources available. All the environmental parameters of concern to the public, scientists, and regulators cannot be measured. A limited set of parameters that serve as indicators of overall estuarine condition needed to be identified, calibrated and verified. It is obvious that one or two samples from a few locations, collected at one time of day, in a single season of a particular year cannot characterize the ecological condition of an estuary. The 1990 Virginian Province Demonstration Project was used to identify which indicators and design attributes are most effective for meeting program objectives and forming the basis for developing programs in other provinces.

The objectives of the 1990 Virginian Province Demonstration Project were to:

- Demonstrate the value of regional monitoring using an unbiased sampling approach as a basis for assessing the condition of estuarine resources;

- Evaluate the ability of a suite of indicators of environmental quality to discriminate between polluted and unpolluted sites over the regional scale;

- Establish standardized methods for monitoring indicators of ecological status and trends in estuaries;

- Obtain data on regional-scale variability in ecological parameters to evaluate and refine the sampling design;

- Develop analytical procedures for using regional-scale monitoring data to assess the ecological status of estuaries, and apply the procedures to establish baseline conditions in the Virginian Province; and

- Identify and resolve logistical problems associated with conducting a regional-scale monitoring program in estuaries.

Clearly, all of these objectives cannot be accomplished in a single year, and most require a process of continuing adjustment and improvement using data collected in subsequent years. This report will describe the current status of analyses for the 1990 Virginian Province Demonstration Project and provide evaluations of the sampling design and the parameters measured during the Demonstration Project.

1.4 PURPOSE AND ORGANIZATION OF THIS REPORT

The purpose of this report is to present results for each of the six objectives of the 1990 Demonstration Project and to evaluate the project itself. Because of its dual purpose as an evaluation of a process and a preliminary product of that process, this report is expected to have many audiences with varying interests and levels of expertise in estuarine science and monitoring; consequently, the degree of detail and description of analyses presented varies among sections.

This report is organized in sections addressing the objectives of the 1990 Demonstration Project. Sections 2 through 5 are evaluations of specific elements of the 1990 Demonstration Project including methods, logistics, indicator development, and sampling design. A description of the 1990 Demonstration Project is presented in Section 2 along with a summary of the standardized sample collection and processing methods and data management procedures developed for monitoring indicators of ecological status and trends in estuaries (Holland 1990; Rosen et al. 1990; Strobel 1990; USEPA 1991). This section also contains a description of the analytical methods used to complete the evaluation of estuarine status. Logistical problems associated with conducting a regional-scale estuarine monitoring program are identified in Section 3. Specific problems relating to quality control are also included in this section (Valente et al. 1990).

Section 4 presents an evaluation of the indicators developed for the Demonstration Project (Holland 1990; Hunsaker and Carpenter 1990; Knapp et al. 1990) and describes a general approach for developing biological indices that can be used to discriminate between polluted and unpolluted sites over regional scales. Data from the Demonstration Project are used in Section 5 to define the regional-scale variability in ecological parameters and evaluate the EMAP sampling design used for conducting a regional-scale monitoring program in estuaries. Additional data from other environmental monitoring programs are used with EMAP data to make preliminary estimates of the power to detect trends using this sampling design.

Sections 2 through 5 use results from the Demonstration Project to build upon material previously presented by the EMAP-Estuaries Resource Group and are intended primarily for technical audiences within EMAP and for reviewers of the program. Section 6 is a preliminary evaluation of the condition of estuarine waters in the Virginian Province in 1990. It is intended primarily for non-technical users of EMAP data, including Congress, the EPA Administrator, the EPA Regions and program offices, and state and local resource managers. The section presents information, distilled from the 1990 Demonstration Project data, that may be of use to regulators and resource managers for determining the scope and utility of EMAP monitoring. It also provides a summary of the information that specialists in estuarine science and monitoring believe to be most pertinent for assessing estuarine condition. The style of Section 6 is purposely different from the other sections to highlight that it is written for a different audience; however, cross-reference is made to later sections to provide technical justifications and explanations as appropriate.

The juxtaposition of the preliminary assessment with the evaluation of the 1990 Demonstration Project affords scientific reviewers and managers alike the opportunity to determine the utility and value of regional monitoring data collected with a probability-based sampling approach, thus addressing one of the primary goals of the 1990 Virginian Province Demonstration Project.

Section 7 summarizes the conclusions that can be drawn from the 1990 Demonstration Project as they relate to its stated objectives. This section is not meant to present major results as does the Executive Summary. Rather, it describes how the results may be of use to potential clients of EMAP.

SECTION 2
DEMONSTRATION PROJECT APPROACH AND METHODS

Three elements of the EMAP approach to monitoring are distinctive and were the guiding forces in designing the sampling plan for the 1990 Demonstration Project. First, probability-based sampling sites are selected in an unbiased manner so that resources are sampled in proportion to their abundance in a size class. Probability-based sampling permits estimation of the condition of the portion of the resource that was not sampled based on knowledge of the sampled portion. Estimates of the proportion of the total area sampled that is in degraded condition can be made with quantifiable confidence, and the level of confidence in the estimate can be increased by increasing the number of sites sampled, if interest and resources allow.

Second, EMAP monitoring focuses on indicators of biological response to stress and uses measures of exposure to stress or contamination as a means for interpreting that response. Traditionally, estuarine monitoring has focused on measures of exposure (e.g., concentration of contaminants in sediment) and attempted to infer biological impacts based upon laboratory bioassays. The advantage of the ecologically-based approach emphasized in EMAP is that it can be applied to situations where multiple stressors exist, acting separately or in combination, and where natural processes cannot be modeled easily. This is certainly the case in estuarine systems, which are subject to an array of anthropogenic inputs and exhibit great biotic diversity and complex physical, chemical, and biological interactions.

Third, EMAP monitoring is conducted on regional and national scales. Standardized methods are employed, and an entire region is sampled simultaneously within a defined time period to ensure comparability of data within and among sampling years. EMAP identified boundaries for eight estuarine regions (Fig. 2-1) based upon biogeographic provinces defined previously by NOAA and the U.S. Fish and Wildlife Service using major climatic zones and prevailing ocean currents (Bailey 1983; Terrell 1979). The 1990 Demonstration Project in the Virginian Province, which includes the wide expanse of irregular coastline from Cape Cod, Massachusetts, to the mouth of the Chesapeake Bay (Cape Henry, Virginia), was designed to evaluate the feasibility of regional sampling and to evaluate and improve the sampling design and indicators for future monitoring. This section summarizes the sampling plan and specific methods used to collect data for a preliminary assessment of the ecological condition of the estuaries of the Virginian Province in 1990. More detailed methods are described by Holland (1990).

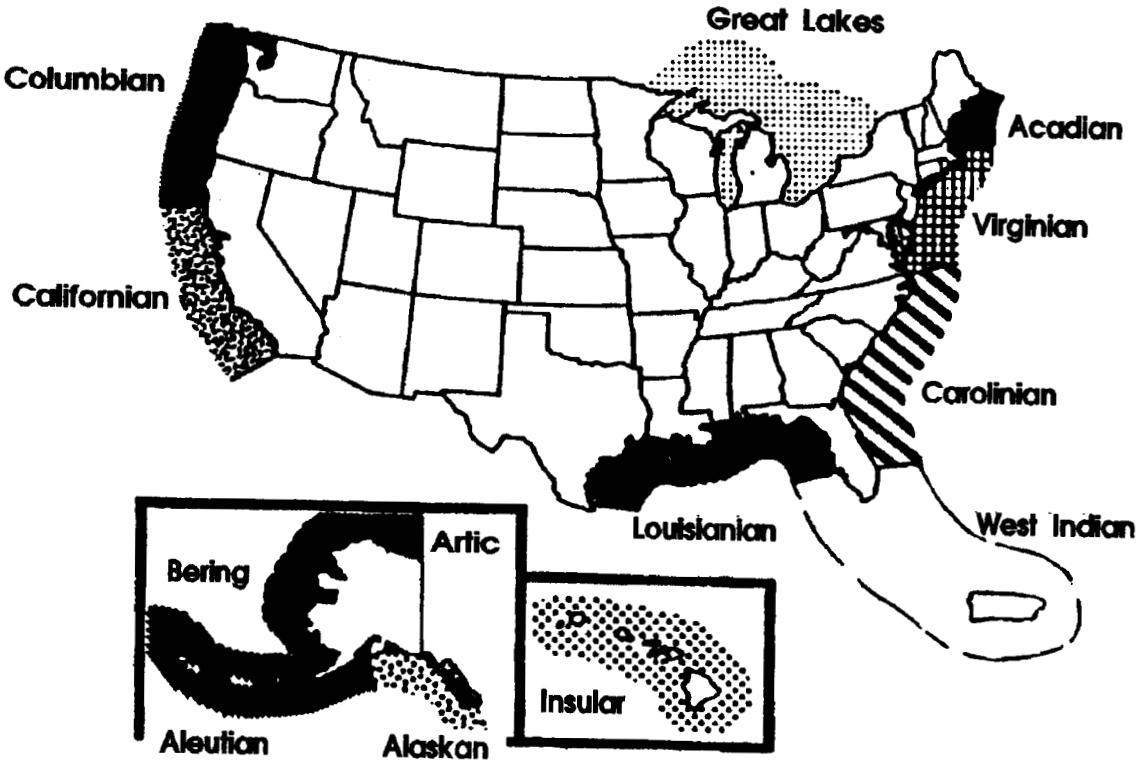

Figure 2-1. EMAP biogeographical provinces for estuaries

2.1 SAMPLING DESIGN

An objective of the 1990 Demonstration Project in the Virginian Province was to collect data at the spatial scale and using the collection methods expected to be used when EMAP monitoring in the Virginian Province is fully implemented. This sampling approach permits evaluation of the logistical feasibility of the sampling plan and generation of preliminary status estimates of the type that will be described in Section 6. The 1990 Demonstration Project also included several pilot elements designed to evaluate the indicators and sampling design. Accomplishment of the pilot study objectives required augmenting the probability-based sampling sites used for developing regional status estimates with judgement samples. This resulted in sampling five types of sites during the Demonstration Project:

- Base sites were the probability-based sampling sites forming the core of the EMAP monitoring design for estuaries. Data collected from these sites are the basis of the preliminary status evaluation for the Virginian Province presented in Section 6.

- Long-term Dissolved Oxygen (LTDO) sites were a subset of the base sample sites. Oxygen meters were deployed to obtain a continuous record of dissolved oxygen concentrations at these sites. These continuous records were used to evaluate the variability of oxygen concentrations and the feasibility of using single, point-in-time measurements of dissolved oxygen to determine status (see Section 4).

- Supplemental sampling sites were part of a pilot study to define an adequate spatial scale for full implementation of EMAP monitoring in estuaries and to define spatial variability in small estuaries. These sites were selected in a probability-based manner and were used in the preliminary status evaluation.

- Indicator testing and evaluation (ITE) sites were part of a pilot study to determine the reliability, sensitivity, specificity, and repeatability of indicator responses for discriminating between polluted and unpolluted conditions. These sites were selected based on historical data and expert judgement to represent the extremes of environmental exposure (degraded to pristine).

- Index sites were part of a pilot study to evaluate whether individual "representative" sites can be used to portray the status of individual small estuaries or tidal river segments in the way that the oxygen content of hypolimnetic water can be used to describe the trophic state of small lakes. Whereas base sites were designed to quantify the areal extent of pollution effects (i.e., percent of degraded estuarine area), index sites were designed to estimate the percent of affected small estuarine systems or tidal river segments.

Potentially, information from base and index sites can be used together to distinguish whether the estimated areal extent of identified problems results from extensive problems in a small number of systems, or from problems affecting a small amount of area in a large number of systems.

Figure 2-2 presents a map of all the probability-based (base and supplemental) sites sampled in the 1990 Demonstration Project. All sites were restricted to a depth greater than 1.5 m to allow for boat access. Therefore, the sampling results may not be applicable to habitats characterized by a depth of less than 1.5m. Methods used to select sampling locations for each type of sampling site are provided below.

2.1.1 Base Sampling Sites

The sampling design for base sites was stratified by size into large estuaries, large tidal rivers, and small estuarine systems. Stratification permitted customizing the sampling frame to the specific geographic features of these different classes of estuaries (Table 2-1). It also allowed allocation of a strata-specific number of samples so that class estimates could be derived with a desired level of precision. The boundaries of these classes were defined using NOAA maps, resulting in 12 large estuaries, 5 large tidal rivers (i.e., Hudson, Potomac, James, Delaware, Rappahannock), and 137 small estuarine systems (Holland 1990). Methods for selecting sampling sites within each stratum are described below.

Large Estuaries – Large estuary sampling sites were selected using an enhancement of the systematic sampling grid proposed for use throughout EMAP (Overton 1989). This grid was placed randomly over a map of the United States and intensified to make 280 km^2 hexagonal grids. Fifty-four base sampling sites were selected using this grid. The sampling sites were the center points of the hexagons, which were 18 km apart.

Large Tidal Rivers – Base sampling sites in large tidal rivers were selected using a "spine" and "rib" approach that is a linear analog of the sampling grid for large estuaries. The starting point of the spine was at the mouth of the river, and the first transect ("rib") was located at a randomly selected river-kilometer between 0 and 25. Additional upstream transects were placed every 25 km from the first. Sampling sites were selected at random along the rib of each transect. A total of 25 base sampling sites were identified in the large tidal rivers of the Virginian Province.

Small Estuarine Systems – A list frame was used to select 32 (23%) of the 137 small estuarine systems in the Virginian Province for sampling during the 1990 Demonstration Project. To ensure that the selected systems were dispersed geographically, all small estuarine systems in the province were listed in order of latitude from north to south and combined into groups of four. One system was selected at random from each group.

Table 2-1. Summary of the characteristics of estuarine classes in the Virginian Province			
Characteristics	**Large Estuaries**	**Large Tidal Rivers**	**Small Estuaries**
Surface Area	> 260 km^2	> 260 km^2	2.6 - 260 km^2
Shape	Aspect ratio < 20	Aspect ratio > 20	Any
Salinity	Strong salinity gradients	Partial salinity gradients	Generally does not have salinity gradients
Watersheds	Large, complex	Large, complex	Small
Management Regions	Multi-state	Multi-state	Usually a single state
Contaminant Sources	Multiple	Multiple	Relatively few
Total area in Virginian Province	16,889 km^2	1,810 km^2	4,879 km^2
Percent of total area in Virginian Province	72	8	21

2.1.2 Long-term Dissolved Oxygen (LTDO) Sites

Thirty of the 116 base sampling sites located throughout the province were selected as LTDO sites. Sites were selected for large estuarine systems (13 sites), large tidal rivers (5 sites), and small estuarine systems (5 sites) and represented a range of estuarine habitats as defined by salinity, sediment type, and depth.

2.1.3 Supplemental Sampling Sites

Available data were insufficient to ascertain the spatial sampling scale necessary to represent the ecological status of estuarine systems in the Virginian Province with adequate precision. To address this problem in large estuaries, the Delaware Bay was sampled at a density four times greater (i.e., sample points were located approximately 9 km apart, yielding 20 additional sampling sites) than other large estuaries. To address the problem in tidal rivers, sampling intensity was doubled in the Delaware River. Supplemental samples were placed

exactly half-way (12.5 km) between base sampling sites there, resulting in four additional samples.

Two types of supplemental sampling sites were selected in small estuaries to determine the appropriate scale for representing resource conditions in this class: 1) an additional randomly-located site was selected in each of five separate systems, and 2) four randomly-located supplemental sites were selected in a single small estuarine system (i.e., Indian River).

2.1.4 Indicator Testing and Evaluation Sites

Based on a review of existing data and expert opinion, 23 indicator testing and evaluation sites were selected for specific combinations of geographic location, salinity, concentrations of sediment contaminants, and dissolved oxygen concentrations. These sites were sampled to investigate the reliability of indicator responses for discriminating between degraded and nondegraded sites across the range of habitats that occurs throughout the Virginian Province (Fig. 2-3).

2.1.5 Index Sites

Index sites were located in deep, muddy, depositional areas, the places believed to be most likely to accumulate contaminants or to experience the most severe stress from low dissolved oxygen concentrations within the sampling period. Index sites were located in every river segment and every small estuary in which there were base sites, resulting in 25 index sites in tidal rivers and 32 in small estuaries.

2.2 SAMPLING METHODS

All sampling was conducted from 8-m (24-ft), twin-engine, Romarine workboats. Specific sampling sites were located using the onboard navigation equipment (Raytheon RAYNAV-780 loran-C) and, when applicable, dead reckoning. Loran-C was used to locate most sites. Dead reckoning was used to locate sites where signal interference prevented reliable performance of loran. Most sites in large tidal rivers and small estuarine systems had obvious landmarks, channel buoys, and other fixed structures that could be used to obtain ranges and bearings.

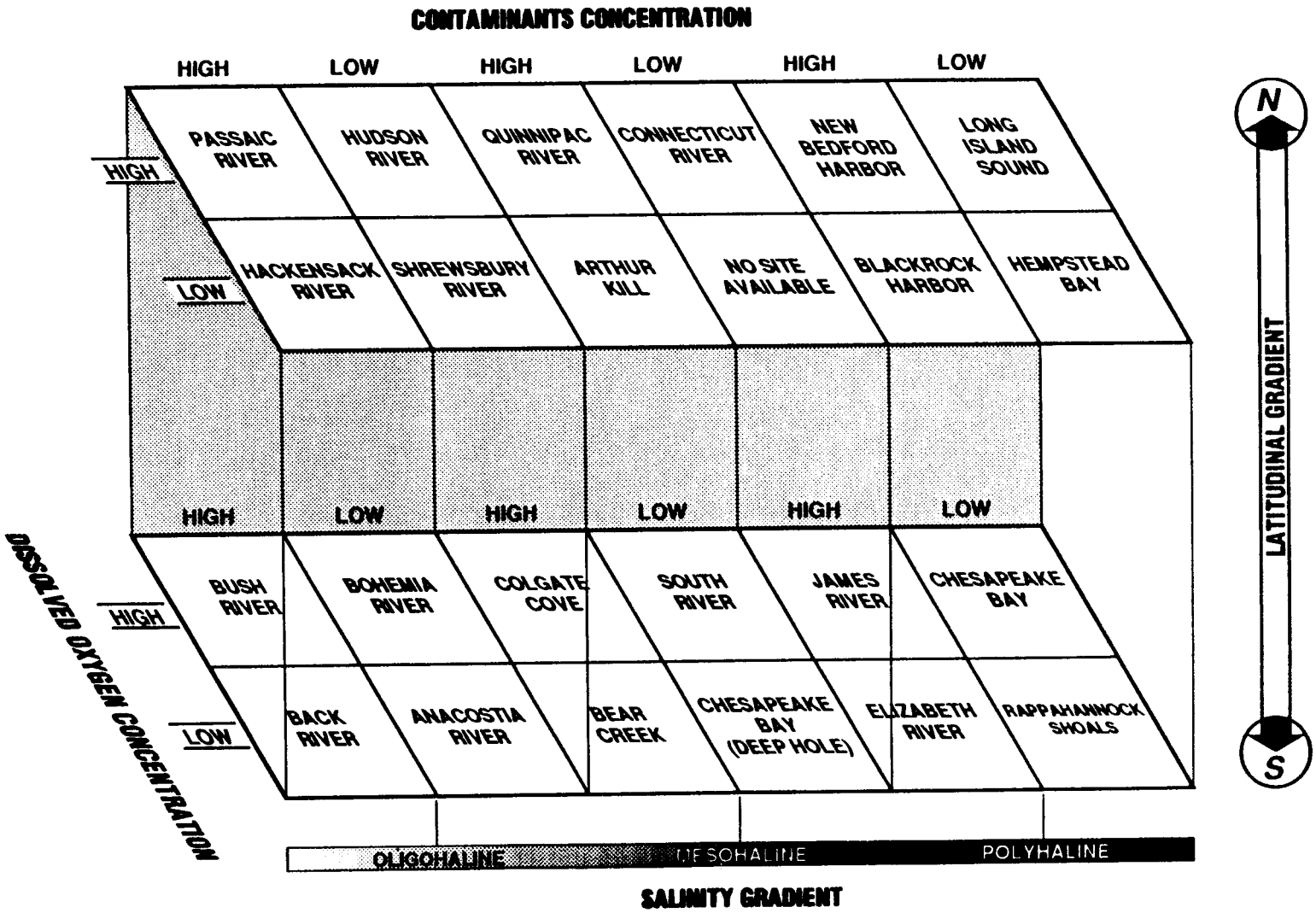

Figure 2-2. Schematic summarizing the indicator testing and evaluation strategy for the 1990 Demonstration Project in the Virginian Province

Sampling was performed by three sampling teams working simultaneously. Team 1 sampled stations from Cape Cod west to the Hudson River. Team 2 sampled from New York Harbor, south to the Delaware Bay, and the northern-most part of the Chesapeake Bay. Team 3 sampled the remaining stations in coastal Delaware, Maryland, and Virginia (Fig. 2-4). Teams circuited their assigned regions every ten days. Stations were organized into clusters, and specific sites were selected at random from each cluster to be sampled on each circuit.

All sampling for the 1990 Demonstration Project was conducted between 19 June and 23 September. The sampling period was divided into three sampling intervals (19 June to 18 July, 19 July to 31 August, 1 September to 23 September). Sampling was repeated in three intervals as a pilot element to determine the stability of various indicators over the summer. Only a subset of parameters was remeasured at all stations; therefore, the number and type of sites sampled, and the types of sampling activities that took place at those sites varied among intervals (Table 2-2). A description of the parameters measured at each site and the specific methods used to conduct the sampling follows.

The EMAP indicator strategy involves four types of ecological indicators (Hunsaker and Carpenter 1990): response, exposure, habitat, and stressor (Fig. 2-5). Response indicators are ecological characteristics that integrate the responses of living resources to specific or multiple pollutants and other stresses and are used by EMAP to assess overall estuarine condition. Exposure indicators quantify pollutant exposure and habitat degradation and are used mainly to identify associations between stresses on the environment and degradation in response indicators. Habitat indicators provide basic information about the natural environmental setting and are used to normalize exposure and response indicators to natural environmental gradients. Stressor indicators are used to quantify pollution inputs or stresses and identify the probable sources of pollution exposure. Examples of the relationships between response, exposure, and habitat indicators sampled during the 1990 Demonstration Project are given in Fig. 2-5. Descriptions were taken from the *Near Coastal Program Plan* (Holland 1990), the *Near Coastal Field Methods Manual* (Strobel 1990), and the *Near Coastal Laboratory Procedures Manual* (USEPA 1991). Readers also are referred to the overall quality assurance plan for the 1990 Demonstration Project (Valente et al. 1990).

2.2.1 Bottom Dwelling (Benthic) Animals

Benthic invertebrate assemblages are composed of diverse taxa with a variety of reproductive modes, feeding guilds, life history characteristics, and physiological tolerances to environmental conditions (Warwick 1980; Frithsen 1989; Bilyard 1987). As a result, benthic populations respond to changes in conditions, both natural and anthropogenic, in a variety of ways (Pearson and Rosenberg 1978; Rhoads et al. 1978; Boesch and Rosenberg

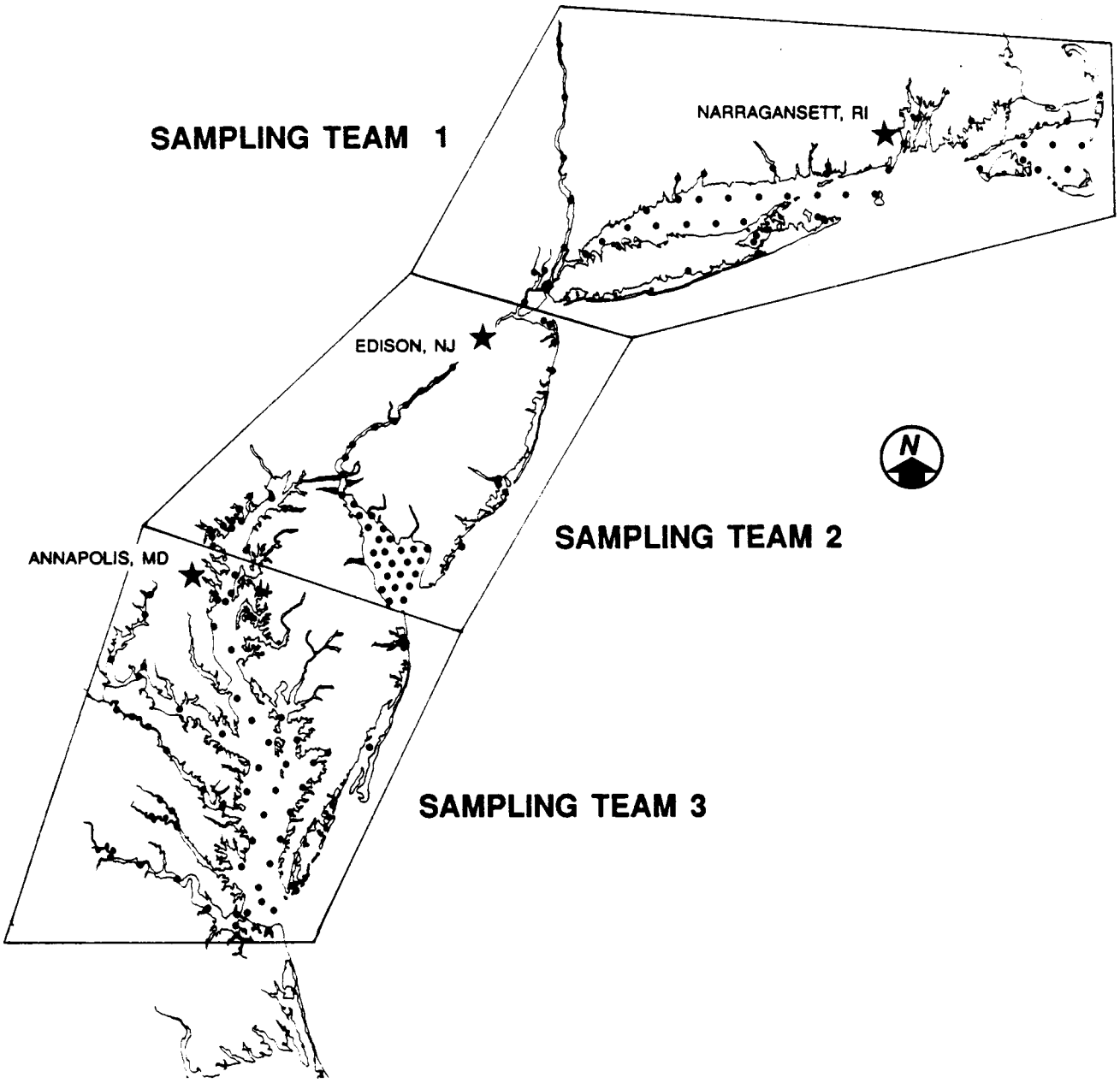

Figure 2-3. Areas sampled by each team during the 1990 Demonstration Project in the Virginian Province. Base Stations are indicated by stars.

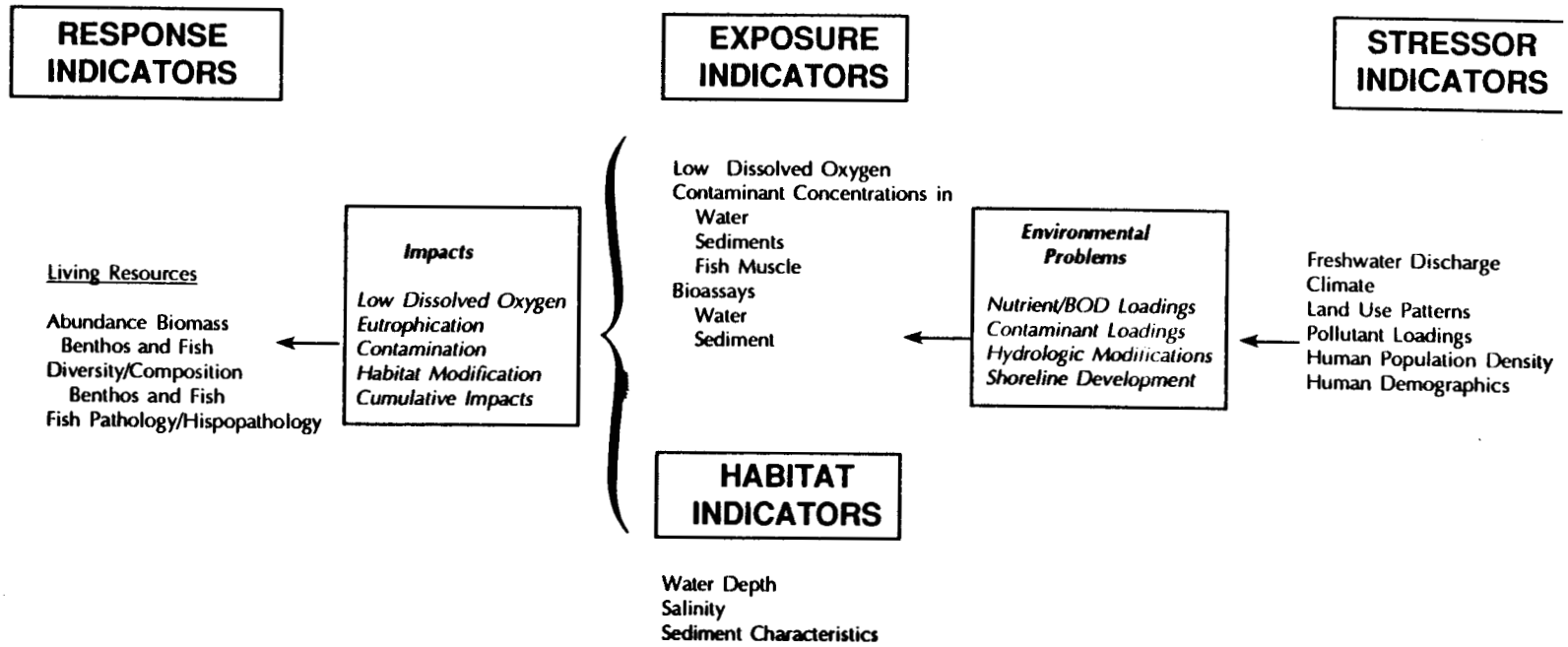

Figure 2-4. Overview of the EMAP indicator strategy giving examples of the types of indicators used to assess estuarine status. The relationships between indicators and the major environmental problems and impacts are shown.

Table 2-2. Sampling activities accomplished at each type of station in each of the three sampling intervals during the 1990 Demonstration Project

Station Type	Interval 1	Interval 2	Interval 3
Base Sampling Site	CTD Profile[a] Fish Trawling[b]	CTD Profile[a] Fish Trawling[b] Benthic Suite[c]	CTD Profile[a] Fish Trawling[b] Shellfish Dredging
Supplement Sites	Not sampled	CTD Profile[a] Fish Trawling[b] Benthic Suite[c]	Not sampled
Continuous Dissolved Oxygen Monitoring Sites	CTD Profile Fish Trawling Benthic Assemblage	CTD Profile[a] Fish Trawling[b] Benthic Suite[c]	CTD Profile Fish Trawling Benthic Assemblage Shellfish Dredging
Index Sites	Not sampled	CTD Profile Benthic Assemblage	Not sampled
Indicator Testing/ Evaluation Sites	Not sampled	CTD Profile[a] Fish Trawling[b] Benthic Suite[c] Bivalve Dredging Water Column Toxicity	Not sampled

(a) CTD profile includes dissolved oxygen, temperature, salinity, pH, PAR, transmissometry, and fluorometry.
(b) Fish trawling includes measurement of fish assemblage, tissue contaminants, and gross pathology.
(c) Benthic suite sampling includes samples for determining benthic invertebrate assemblage, sediment contaminants, and sediment toxicity.

1981). Responses of some species (e.g., filter feeders, species with pelagic life stages) indicate changes in water quality and others (e.g., organisms that burrow in or feed on sediments) changes in sediment quality; furthermore, because most benthic species have limited mobility, they cannot avoid exposure to pollution stress. Benthic community studies have been used in many regional estuarine monitoring programs (Bilyard 1987; Holland et al. 1987) and have proven to be an effective indicator for describing the extent and magnitude of pollution impacts in estuarine ecosystems, as well as for assessing the effectiveness of management actions.

Benthic samples for measures of species composition, abundance, and biomass were collected at all sampling sites during the second interval and at continuous monitoring sites during the first and third intervals. Samples were collected with a Young-modified Van Veen grab that samples a surface area of 440 cm^2; three grabs were collected at each site. The Young grab was selected because it is deployed easily from small boats, and it samples both mud and sand habitats adequately. The maximum depth of penetration of the grab was 10 cm, and only grabs that penetrated deeper than 7 cm were accepted. A small core was taken from each grab to determine silt-clay content. The remaining sample was sieved through a 0.5 mm screen using a backwash technique that minimized damage to soft-bodied biota. Samples were preserved in 10% buffered formaldehyde-rose bengal solution and stored for at least 60 days prior to processing.

In the laboratory, macrobenthos were sorted, identified to the lowest practical taxonomic level, and counted. All macrobenthos were identified to species, except for the following groups:

Taxonomic Group	Level of Identification
Class Anthozoa	Class
Subclass Copepoda	Order
Phylum Nemertinea	Phylum
Subclass Ostracoda	Subclass
Class Turbellaria	Class

For samples collected in low salinity (less than 5 ppt) water, oligochaetes were identified to species, and chironomids to genus. Above 5 ppt salinity, oligochaetes were identified to class (most were assumed to belong to the genus Tubificoides), and chironomids were identified only to family. Biomass was measured for approximately 30 species that were expected, prior to the beginning of the 1990 Demonstration Project, to be dominant (Table 3-3). All other species were combined into biomass groups defined by feeding type and major taxonomic group (i.e., sub-surface, deposit-feeding polychaetes). Bivalves and polychaetes were classified into feeding categories defined by Fauchald and Jumars (1979), Jorgenson (1966), and Bousfield (1973). Shell-free dry weight after drying at 60°C was determined using an analytical balance with an accuracy of 0.1 mg. Large bivalves (greater than 2-cm long) were shucked prior to determining biomass. Smaller shells were removed by acidification using a 10% HCl solution.

Table 2-3. Biomass groups for macrofaunal species. Biomass determined for each species or group of species listed.

Amphipods		
Ampelisca sp.	Haustoriidae	*Unciola* sp.
Corophium sp.	*Leptocheirus* sp.	Other Amphipods
Gammaridae	*Monoculodes* sp.	
Isopods		
Cyathura sp.		Other Isopods
Decapods		
All		
Chironomids		
Intact individuals and anterior fragments		
Posterior fragments		
Polychaetes		
Glycera sp.	*Marenzelleria viridis*	*Paraprionospio pinnata*
Heteromastus filiformis	*Mediomastus ambiseta*	*Polydora* sp.
Leitoscoloplos sp.	*Neanthes succinea*	*Streblospio benedicti*
Maldanidae	*Nephtys* sp.	
Other sub-surface deposit feeding polychaetes		Other carnivore/omnivore polychaetes
Other surface feeding polychaetes		Unidentified polychaetes and fragments
Oligochaetes		
Intact individuals and anterior fragments		
Posterior fragments		
Bivalves		
Corbicula fluminea	*Mulinia lateralis*	*Yoldia limatula*
Ensis directus	*Nucula* sp.	Other deposit feeding bivalves
Gemma gemma	*Rangia cuneata*	Other suspension feeding bivalves
Mercenaria mercenaria	Tellinidae	Unidentified bivalves
Gastropods		
Acteocina canaliculata		
Hydrobia sp.		
Other Gastropods		
Miscellaneous		
Echinodermata		Nemertinea
Hemichordata		
Phoronis sp.		

2.2.2 Sediment Profile Images

Benthic animals are important regulators of the deposition and resuspension of bottom sediments, the exchange of constituents between bottom sediments, and the exchange of constituents between bottom sediments and the overlying water (Rhoads 1974; Rhoads et al. 1978; Rhoads and Boyer 1982; Aller 1982). By ventilating and displacing sediments during burrowing and feeding, they affect geochemical profiles in sediments and pore waters. This is particularly true in higher salinity habitats with fine-grained sediments, where there is no wave disturbance or tidal scour. In these habitats, the apparent depth to the redox potential discontinuity (RPD), the oxygenated layer of sediments, is positively associated with the presence of larger, deep-burrowing, longer-lived organisms, and inversely related to the presence of smaller, short-lived, surface dwelling, opportunistic species. The former condition is representative of acceptable ecological conditions, and the latter may be associated with physical or anthropogenically induced disturbance. Sediments exhibiting a shallow RPD that are dominated by shallow-burrowing, short-lived species have been shown to be either chemically or organically enriched (Rhoads and Germano 1986; Scott et al. 1987; O'Connor et al. 1989; Valente et al. 1992).

Sediment profile images were collected at 20 stations in the northern half of the Virginian Province. Replicate photographic images were taken using a Benthos Model 3771 sediment profile camera and the Remote Ecological Monitoring of the Seafloor (REMOTS)® imaging analysis system (Rhoads and Germano 1982, 1986). The camera photographs the sediment-water interface in the vertical plane. The photograph is processed to quantify apparent RPD depth, grain size, relative abundance of surface tube structures, boundary roughness, penetration depth, and the presence of feeding voids and methane gas bubbles. The depth of the apparent RPD is determined in the sediment profile images as the boundary between light-colored, oxidized surface sediments and underlying grey to black reduced sediments. The successional stage is characterized by the types of fauna present, or whose presence is inferred, in the REMOTS® image. Stage I communities typically consist of small tubicolous spionid or capitellid polychaetes that exploit recently disturbed or open space. Stage II represents a transitional community that commonly consist of dense aggregations of tubicolous amphipods (e.g., *Ampelisca*) and tellinid bivalves. Stage III fauna are represented by relatively large, head-down deposit feeders. These are rarely observed directly on the images, but their presence is inferred by feeding pockets or voids at depth. Chemical parameters are indicated by the presence of highly reflective methane gas bubbles in the sediment, and low reflectance (sulfitic) sediment near the sediment-water interface (high sediment oxygen demand). A multi-parameter index, the Organism-Sediment Index (OSI), which ranges from -10 to +11, was calculated to characterize sediment quality (Table 2-4). The OSI is based on measurements of the RPD depth, infaunal successional stage, and the apparent presence of selected chemical attributes (e.g., methane gas).

Table 2-4. Calculation of the OSI based on sediment profile images	
Mean Apparent RPD Depth Measurement	**Index Score**
0.00 cm	0
>0 to 0.75 cm	1
0.76 to 1.50 cm	2
1.51 to 2.25 cm	3
2.26 to 3.00 cm	4
3.01 to 3.75 cm	5
>3.75 cm	6
Infaunal Successional Stage	
No visible macrofauna	-4
Small, short-lived, surface-dwelling opportunists: Stage I	1
Transitional, surface-dwelling, tubicolous amphipods, tellinid bivalves: Stage II	3
Large, head-down, deposit feeding polychaetes and bivalves: Stage III	5
Stage I on III	5
Stage II on III	5
Chemical Parameters	
Methane gas present	-2
High apparent sediment oxygen demand	-4
OSI = Total of chosen index values (range: -10 to +11)	

2.2.3 Fish

There are several advantages to using fish as potential indicators of estuarine condition. Because of their longevity and dominant position at the upper end of the food web, fish responses integrate many short-term and small-scale environmental perturbations. They are

known to respond to most of the major environmental problems of concern in estuaries, including eutrophication, habitat modification, and pathogenic or toxic contamination. Eutrophication can affect fish adversely by reducing dissolved oxygen below levels that are critical for growth or survival. Habitat modification, such as the loss of submerged aquatic vegetation, has been linked to decreased fish productivity through loss of important nursery areas. Toxic and pathogenic contaminants can decrease fish growth, reproduction, or survival and can make fish unsafe for human consumption. Fish also are valuable as indicators because of their importance for determining the public perception of estuarine quality. Reports of fishery closures due to chemical and viral contamination alarm the public.

Fish were collected by trawling with a 16-m (footrope), high-rise net with 5-cm mesh wings and a 2.5-cm mesh cod end. The net was towed for 10 minutes against the tide at speeds between 0.6 and 1.0 m/s (i.e., speed over the bottom between 0.3 and 1.0 m/s). All fish caught in a trawl were identified to species and counted; up to 30 fish of each species from each collection were measured to the nearest millimeter. Up to five individuals in predetermined size ranges from each of 10 target species (Table 2-5) were retained from each trawl for tissue analysis. The specimens were gutted, labeled, frozen on dry ice, and shipped to the laboratory, where they were stored frozen to await analysis for the same chemicals measured in the NOAA National Status and Trends Program (Table 2-6). No PAHs were planned for tissue residue analyses because of the low probability of finding PAHs in muscle tissue. Delays in the laboratory resulted in exceeding the holding time for frozen samples; consequently, fish tissue samples were not processed as part of the 1990 Demonstration Project.

Table 2-5. Targeted species and size ranges of fish retained for tissue analysis during the 1990 Virginian Province Demonstration Project. If fish from the primary size range were not captured, fish were selected from the secondary range.

Target Species	Primary Size Range (mm)	Secondary Range (mm)
Channel Catfish (*Ictalurus punctatus*)	200-300	300-400
Atlantic Croaker (*Micropogon undulatus*)	200-300	300-400
Hogchoker (*Trinectes maculatus*)	100-150	150-250
Summer Flounder (*Paralichthys dentatus*)	350-450	200-350
Spot (*Leiostomus xanthurus*)	150-250	100-150
White Catfish (*Ictalurus catus*)	200-300	300-400
Weakfish (*Cynoscion regalis*)	300-400	200-300
Winter Flounder (*Pseudopleuronectes americanus*)	300-400	200-300
Windowpane Flounder (*Scophthalmus aguosus*)	300-400	200-300
White Perch (*Morone americana*)	150-250	100-150

Table 2-6. Analytical measurements for fish and bivalve tissue samples collected during the 1990 Virginian Province Demonstration Project

DDT and its Metabolites	Trace Elements	
o,p'-DDD	Aluminum	Mercury
p,p'-DDD	Arsenic	Nickel
o,p'-DDE	Cadmium	Selenium
p,p'-DDE	Chromium	Silver
o,p'-DDT	Copper	Tin
p,p'-DDT	Iron	Zinc
	Lead	

Chlorinated Pesticides other than DDT	
Aldrin	Heptachlor epoxide
Alpha-Chlordane	Hexachlorobenzene
Trans-Nonachlor	Lindane (gamma-BHC)
Dieldrin	Mirex
Heptachlor	

18 PCB Congeners:	
PCB No.	**Compound Name**
8	2,4'-dichlorobiphenyl
18	2,2',5-trichlorobiphenyl
28	2,4,4'-trichlorobiphenyl
44	2,2',3,5'-tetrachlorobiphenyl
52	2,2',5,5'-tetrachlorobiphenyl
66	2,3',4,4'-tetrachlorobiphenyl
101	2,2',4,5,5'-pentachlorobiphenyl
105	2,3,3',4,4'-pentachlorobiphenyl
118	2,3',4,4',5-pentachlorobiphenyl
128	2,2',3,3',4,4'-hexachlorobiphenyl
138	2,3',3,4,4',5-hexachlorobiphenyl
153	2,2',3,4,4',5'-hexachlorobiphenyl
170	2,2',4,4',5,5'-hexachlorobiphenyl
180	2,2',3,3',4,4',5-heptachlorobiphenyl
187	2,2',3,4,4',5,5'-heptachlorobiphenyl
195	2,2',3,3',4,4',5,6-octachlorobiphenyl
206	2,2',3,3',4,4',5,5',6-nonachlorobiphenyl
209	decachlorobiphenyl

At all stations where fish were collected, up to 30 individuals of each target species were inspected for visible external pathological disorders. This inspection included checking the

body surface and fins for skin discoloration, raised scales, white or black spots, ulcers, fin erosion, lumps or growths, parasites, and opercular deformity; the branchial chamber for gill discoloration, erosion, deformity, parasites, and lumps or growths; the buccal cavity for hemorrhages, parasites, and lumps or growths; the overall morphology of the fish for skeletal malformations; and the condition of the eyes (e.g., cloudiness, hemorrhages, enlarged or sunken). Specimens with pathologies were preserved in Dietrich's solution for laboratory verification. At ITE sites, up to 30 specimens of each target species and 20 specimens of non-target species that were free of visible pathological abnormalities were preserved for quality control checks of field observations. At indicator testing and evaluation sites, tissue samples for histological examination were collected from all target fish that exhibited visible pathology. For comparison, tissue samples were taken from up to 25 target fish and 10 non-target fish that did not exhibit visible pathology. Tissue samples were dehydrated in an ethanol gradient, cleared in a xylene substitute, infiltrated, and embedded in paraffin. Sections were cut at 6 μm on a rotary microtome, stained with Harris' hematoxylin and eosin, and examined microscopically. These examinations were intended to determine the relationship between the incidence of visible external abnormalities and the presence and kinds of internal histopathological changes, and to evaluate the ability of this indicator to discriminate between polluted and unpolluted sites.

2.2.4 Large Bivalve Abundance and Tissue Contamination

The occurrence of large, older bivalves at a site generally indicates that environmental conditions have been relatively stable over time. The relative immobility of bivalves makes them good integrators of long-term environmental conditions at the site from which they were collected. The burrowing habit of many bivalves places them where exposure to stress, such as low dissolved oxygen concentrations and contaminants, is likely to be high. Filter feeding bivalves concentrate contaminants ingested with food to levels many times higher than those in the water. Tissue contamination can reduce growth and survival, which can adversely affect production. It also reduces the value and quality of bivalve meats for human consumption.

Attempts were made to collect large infaunal bivalves from ITE sites and from base sampling sites during interval 3 to determine if a sufficient number of specimens could be collected to evaluate the magnitude and primary sources of variation in the concentrations of contaminants in bivalve tissue. If most of the natural variation in bivalve tissue contamination can be accounted for, then contamination in large bivalves may be a valuable indicator of suitability for human use. A 30-cm rocking chair dredge was used to collect specimens of large infaunal species. The dredge, equipped with a 2.5-cm mesh cod-end bag, was towed over the bottom for five minutes at approximately one knot. Mollusks were identified to species and counted. Shell length was measured to provide an indication of the age structure of the population. Up to 10 individuals of each species were labeled, frozen on dry ice, and shipped to the analytical

laboratory, where they were stored frozen to await analysis. Bivalve tissue samples were analyzed using standard methods to determine the concentrations of the contaminants listed in Table 2-6.

2.2.5 Dissolved Oxygen Concentration

Dissolved oxygen (DO) concentration is an important indicator of environmental condition because it is a fundamental requirement for maintenance of populations of benthos, fish, shellfish, and other aquatic biota. DO concentrations are affected by environmental stresses, such as point and nonpoint sources of nutrients, particulates, and dissolved organic matter. Stresses that occur in conjunction with low DO concentrations may be even more detrimental to biota (e.g., exposure to hydrogen sulfide, decreased resistance to disease and contaminants). DO levels, however, are highly variable over time, fluctuating widely due to the tide, winds, and biological activity (Kemp and Boynton 1980; Sanford et al. 1990; Welsh and Eller 1991). The objective of the 1990 Demonstration Project was to collect regional data to determine the best measure for representing dissolved oxygen exposure (e.g., percent of time below a critical value, mean over a defined time period), to evaluate the stability of DO levels over the sampling period, and to determine if the occurrence, magnitude, and duration of exposure to extreme low dissolved oxygen stress can be predicted using short-term continuous records of dissolved oxygen concentration.

Dissolved oxygen was sampled in two ways during the 1990 Demonstration Project: 1) point-in-time water column profiles, and 2) continuous bottom measurements. Point-in-time measurements were made at all sampling sites during every visit. Data from the point measurements were used to generate regional status estimates. Continuous monitoring was conducted at 23 base sampling sites selected to represent a range of geographic and habitat types in the Virginian Province. These sites were known as long-term dissolved oxygen (LTDO) sites. Continuous monitoring data were used in Monte Carlo simulations to evaluate alternative sampling strategies for measuring dissolved oxygen cost-effectively (Section 4).

A Seabird model SBE 25 CTD equipped with a Beckman polarographic DO electrode was used to make the point-in-time measurements. A CTD cast was performed during all station visits in each of the three sampling intervals to obtain vertical profiles of the water column. In addition to producing profiles of dissolved oxygen, the CTD measured salinity, temperature, pH, transmissometry, fluorometry, and photosynthetically active radiation (PAR). At each site, the CTD was held at the water surface for approximately one minute to allow the DO probe to reach thermal equilibrium. Upon reaching equilibrium, the CTD was lowered through the water column at a rate of approximately 1 m per second, until it reached 1 m from the bottom. The CTD was held at this position for two minutes, then retrieved and downloaded to an onboard laptop computer.

Hydrolab DataSonde 3 data loggers were used at the 23 continuous-monitoring sites. These instruments were programmed to record DO measurements every 30 minutes and were suspended approximately 1 m from the bottom at each continuous-monitoring site for 70 days between 19 June and 30 August (sampling intervals 1 and 2). The Hydrolab was retrieved approximately every 10 days. The stored data was downloaded to a computer, and a calibrated Hydrolab was redeployed.

2.2.6 Chlorophyll *a*

The concentration of chlorophyll *a* in the water column is a measure of the biomass of phytoplankton communities (Parsons et al. 1977). This measurement was included in the 1990 Demonstration Project as a potential indicator of eutrophication, since phytoplankton biomass is expected to increase in eutrophic areas receiving large inputs of nutrients.

A Go-Flo sample bottle was used to collect water samples from 1 m below the surface. A portion of the water sample was filtered through a 25-mm glass fiber filter (type GF-F, 0.7-µm nominal pore size) using a 60-ml syringe filter assembly. The filter was removed from the syringe assembly using forceps, wrapped in aluminum foil, placed in a small zip-lock bag, and frozen on dry ice. Samples were kept frozen until analysis.

2.2.7 Sediment Toxicity

Sediment toxicity testing is the most direct measure available for determining the toxicity of contaminants in sediments to indigenous biota. It improves upon direct measurement of sediment contamination because many contaminants are tightly bound to sediment particles or are chemically complexed and, therefore, are not biologically available (USEPA 1989). Sediment toxicity testing cannot be used to replace direct measurement of the concentrations of contaminants in sediment because such measurements are an important part of interpreting the results of toxicity tests.

Sediment for toxicity tests was collected with the Young-modified Van Veen grab used for benthic invertebrate sampling. Grabs were conducted only during the second sampling interval at all but index sites. The top 2 cm of sediment from five or more grabs were removed and placed into a teflon bowl. Care was taken to avoid collecting material from the edge of grabs and to use only samples that had undisturbed sediment surfaces. The teflon bowl was kept on ice in a cooler between grabs to reduce the temperature of the sample and to prevent accidental contamination. After approximately 3,000 ml of sample was collected, the composite was homogenized and distributed to appropriate sediment chemistry, toxicity, and grain size sample containers.

Toxicity tests were performed on the composites of surface sediments collected from each station. Tests were conducted using the standard 10-day acute test method (Swartz et al.

1985; ASTM 1990) and the tube-dwelling amphipod *Ampelisca abdita.* Five replicate tests were completed. For each toxicity test, 200 ml of sediment from the composited sample was placed in 1-l glass jars and covered with 600 ml of 30 ppt seawater. The bioassays were conducted under static conditions for 10 days at 20°C. An additional series of toxicity tests was conducted to determine whether contaminated sediments from low salinity waters became less toxic in the higher salinity waters (30 ppt) used for these tests; thus, for all samples from stations with bottom salinities of 5 ppt or less, a second 10-day acute test was conducted concurrently using the freshwater amphipod, *Hyalella azteca.* The procedure for both species was identical, except that the low salinity test with *Hyalella* used well water.

2.2.8 Sediment Contaminant Concentrations

Metals, organic chemicals, and fine-grained sediments entering estuaries from freshwater inflows, point sources of pollution, and various nonpoint sources including atmospheric deposition, generally are retained within estuaries and accumulate in the sediments (Turekian 1977; Forstner and Wittman 1981; Schubel and Carter 1984; Nixon et al. 1986; Hinga 1988). This is because most contaminants have an affinity for adsorption onto particles (Hinga 1988; Honeyman and Santsche 1988). Chemical and microbial contaminants generally adsorb to fine-grained materials in the water and are deposited on the bottom, accumulating at deposition sites, including regions of low current velocity, deep basins, and the zone of maximum turbidity. The concentration of contaminants in sediments is dependent upon interactions between natural (e.g., physical sediment characteristics) and anthropogenic factors (e.g., type and volume of contaminant loadings; Sharp et al. 1984).

Sediment samples for contaminant analysis were collected during the second sampling interval at all sites except index sites. Samples were collected as a subsample of the sediment slurry collected for sediment toxicity testing. The sediment was placed in clean glass jars with teflon lid liners, shipped on ice, and stored frozen in the laboratory prior to analysis. Sediments were analyzed for the NOAA National Status and Trends suite of contaminants for sediments (Table 2-7) using standard analytical methods (Table 2-8).

2.2.9 Sediment *Clostridium* Spores

Clostridium perfringens is an obligate-anaerobic, spore-forming bacterium found in the feces of warm-blooded animals. Spores accumulate in sediments and have been interpreted as conservative tracers of fecal contamination because the spores survive longer than other common indicators of fecal pollution (Bisson and Cabelli 1980; Duncanson et al. 1986).

Table 2-7. Analytical measurements for sediment samples collected during the 1990 Virginian Province Demonstration Project

Polyaromatic Hydrocarbons (PAHs)

Acenaphthene	2,6-dimethylnaphthalene	Perylene
Anthracene	Fluoranthene	Phenanthrene
Benz(a)anthracene	Fluorene	Pyrene
Benzo(a)pyrene	Ideno(1,2,3-c,d)pyrene	Benzo(b)fluoranthene
Benzo(e)pyrene	2-methylnaphthalene	Acenaphthylene
Biphenyl	1-methylnaphthalene	Benzo(k)fluoranthene
Chrysene	1-methylphenanthrene	Benzo(g,h,i)perylene
Dibenz(a,b)anthracene	Naphthalene	2,3,5-Trimethylnaphthalene

DDT and its metabolites		Chlorinated pesticides other than DDT	
o,p'-DDD	p,p'-DDE	Aldrin	Heptachlor epoxide
p,p'-DDD	o,p'-DDT	Alpha-Chlordane	Hexachlorobenzene
o,p'-DDE	p,p'-DDT	Trans-Nonachlor	Lindane (gamma-BHC)
		Dieldrin	Mirex
		Heptachlor	

Major Elements	Trace Elements		
Aluminum	Antimony	Copper	Selenium
Iron	Arsenic	Lead	Silver
Manganese	Cadmium	Mercury	Tin
	Chromium	Nickel	Zinc

18 PCB Congeners:

No.	Compound Name
8	2,4'-dichlorobiphenyl
18	2,2',5-trichlorobiphenyl
28	2,4,4'-trichlorobiphenyl
44	2,2',3,5'-tetrachlorobiphenyl
52	2,2',5,5'-tetrachlorobiphenyl
66	2,3',4,4'-tetrachlorobiphenyl
101	2,2',4,5,5'-pentachlorobiphenyl
105	2,3,3',4,4'-pentachlorobiphenyl
118	2,3',4,4',5-pentachlorobiphenyl
128	2,2',3,3',4,4'-hexachlorobiphenyl
138	2,3',3,4,4',5-hexachlorobiphenyl
153	2,2',3,4,4',5'-hexachlorobiphenyl
170	2,2',4,4',5,5'-hexachlorobiphenyl
180	2,2',3,3',4,4',5-heptachlorobiphenyl
187	2,2',3,4,4',5,5'-heptachlorobiphenyl
195	2,2',3,3',4,4',5,6-octachlorobiphenyl
206	2,2',3,3',4,4',5,5',6-nonachlorobiphenyl
209	decachlorobiphenyl

Other measurements

Tributyltin	Acid volatile sulfides	Total organic carbon	Grain size distribution

Table 2-8. Analytical methods used in 1990 for determination of chemical contaminant concentrations in sediments	
Compound(s)	**Method**
Inorganics:	
Ag, Al, Cr, Cu, Fe, Mn, Ni, Pb, Zn	Total digestion using HF/HNO_3 (open vessel hot plate) followed by inductively coupled plasma-atomic emission spectrometry (ICP-AES) analysis.
As, Cd, Sb, Se, Sn	Microwave digestion using HNO_3/HCl followed by graphite furnace atomic absorption (GFAA) analysis.
Hg	Cold vapor atomic absorption spectrometry
Organics:	
Extraction/Cleanup	Soxhlet extraction, extract drying using sodium sulfate, extract concentration using Kuderna-Danish apparatus, removal of elemental sulfur with activated copper, removal of organic interferents with GPC and/or alumina.
PAH measurement	Gas chromatography/mass spectrometry (GC/MS)
PCB/pesticide	Gas chromatography/electron capture detection (GC/ECD) with second column confirmation

The concentration of *Clostridium* spores was measured from subsamples of the homogenized sediment samples collected for toxicity tests and contaminant measurements. Spores were extracted from sediment samples by vortexing sediment slurries following the addition of deionized water and sodium metaphosphate. Extracted water was filtered through sterile membrane filters, which were cultured on agar (Emerson and Cabelli 1982). Colonies were counted after staining with ammonium hydroxide.

2.2.10 Sediment Grain Size

The physical characteristics of estuarine sediments (e.g., grain size, silt-clay content) influence the distribution of benthic fauna and the accumulation of contaminants in sediments (Rhoads 1974; Plumb 1981). Sediment grain size and silt-clay content data were collected to help interpret the responses of these parameters. A subsample from the benthic invertebrate and sediment contaminant grabs was retained to determine the silt-clay content and the

distribution of grain sizes in the sediment. Samples for the determination of silt-clay content and grain size distribution were sieved through 63-µm mesh. Both the filtrate and the fraction retained on the sieve were dried in an oven at 60°C and weighed to calculate the proportion of silts and clays in the sample. Procedures for determining grain size distribution generally followed the framework described for the silt-clay analyses; however, the fraction retained on the sieve was processed through an additional sieve analysis to determine specific grain size percentages. The filtrate was processed using pipette analysis.

2.2.11 Water Column Toxicity

Water column toxicity tests were used as indicators of the presence of contaminant concentrations potentially toxic to planktonic estuarine organisms. Samples for water column toxicity tests were collected from the ITE sites during the second sampling interval. A 2-l water sample was taken from 1 m below the surface using a clean Go-Flo sampling bottle. The water sample was placed on ice and shipped immediately to the laboratory so that the tests could be performed within 48 hours of sample collection.

Three water column bioassays were conducted: 1) sea urchin fertilization test (Nacci et al. 1987), 2) red algae sexual reproduction test (Thursby and Steel 1987), and 3) bivalve fertilization and larval growth test (APHA 1985). The results of the three tests were compared to determine their relative sensitivity and were correlated with the results of other indicators to determine whether this indicator should be included in future EMAP monitoring in estuaries.

2.2.12 Marine Debris

Aesthetic appeal is an important factor determining the suitability of an estuary for human use. The presence of trash in the water and the clarity of the water are the primary means by which the public assesses the aesthetic quality of an estuary. Observations of these factors were made at all sites during the 1990 Demonstration Project because of their importance to the public and the relatively small incremental cost of the effort. The kinds and relative amounts of floating and submerged (i.e., collected in otter trawls) trash were noted for all stations. Trash was categorized as paper and plastics, cans and bottles, medical, and other wastes.

2.3 DATA COLLECTION AND SAMPLE TRACKING

Field crews were supplied with two personal computers and appropriate software to facilitate collection and electronic recording of data, and sample tracking. The software included a navigation system linked to loran-C and Global Positioning System (GPS) navigation aids. The navigation system recorded the location and time of day for each observation and was used to identify boat speed and distance for all fish trawls.

All samples, data diskettes, shipments, and equipment were bar coded to facilitate sample tracking and to reduce transcription errors. Field computers were equipped with bar code readers to record sample identification information. The field computer system included software to capture sample identifications, visual observations made by field crews, and all electronic data from the CTD profiles and the continuous dissolved oxygen meter.

Copies of all data were made on diskettes at the end of each day to minimize data loss from potential equipment failure. With the exception of the CTD and continuous dissolved oxygen meter data, records also were kept on paper data sheets. Communications software and commercial carrier phone lines were used to transfer data daily to the Field Operations Center at the USEPA Environmental Research Laboratory in Narragansett, RI. All electronic records were verified with paper forms.

The bar-coded information was used to follow all sample shipments from point-of-origin to point-of-analysis. Samples were followed through processing and analysis steps using an individual sample identification number. Further details on data management for the Demonstration Project are presented in Rosen et al. (1990).

2.4 ANALYTICAL METHODS FOR STATUS ASSESSMENT

Three types of analyses were conducted for this report, including those to 1) test and develop indicators, 2) evaluate the overall sampling design, and 3) conduct an initial evaluation of the status of estuaries within the Virginian Province. The analytical steps used to develop indicators and evaluate the sampling design and the results of those analyses are described in Sections 5 and 6 of this report. The analytical methods used to complete the evaluation of estuarine status that appeared in Section 2 are given below in three parts. The first describes the analytical procedures used to translate measurements made in the field and data generated in the laboratory into information for estimating estuarine status. The second describes the general analytical methods used to generate areal estimates of estuarine condition, and the third contains methods for calculating confidence intervals associated with the areal estimates.

2.4.1 Assessment Methods for Individual Indicators

2.4.1.1 Benthic Index

The benthic assemblage at each sampling site was characterized as degraded or nondegraded based on a benthic index described in Section 4. Degraded and nondegraded conditions were defined relative to regional reference conditions identified from indicator testing stations (Section 4). The index was developed from a discriminant function that incorporates five characteristics of the benthic assemblage: proportion (expressed as a percent) of the expected number of species adjusted for salinity, number of amphipods, percent of total abundance in molluscan taxa, number of capitellids, and average weight per individual polychaete. Only benthic infaunal species were included in this analysis and all data were transformed [Log_{10} (value +1)] prior to conducting the analysis.

The discriminant function combining the five indicators was:

Discriminant Score =

(0.011 * Proportion of expected number of species) +
(0.817 * Number of amphipods) +
(0.671 * Percent of total abundance as bivalves) +
(0.465 * Number of capitellids) +
(0.577 * Average weight per individual polychaete)

Prior to calculating discriminant scores, the log transformed values for each indicator were normalized using a standard normal (Z) transformation:

$$Z = \frac{\chi - \mu}{\sigma}$$

where **Z** is the transformed indicator value and χ is the log of the indicator. The standardized means and standard deviations for each indicator were calculated using the indicator testing stations and are given in Table 2-9.

Table 2-9. Standardized mean (μ) and standard deviation (σ) for indicators used in the discriminant function for the benthic index. All values were calculated with data from the indicator testing stations only.

Indicator	Transformed Mean (μ)	Transformed Standard Deviation (σ)
Proportion of affected number of species	1.4503	0.5498
Number of Amphipids	0.7986	0.8460
Percent of total abundance as bivalves	0.4647	0.5422
Number of Capitellids	0.7187	1.0031
Average weight per individual polychaete	0.0003967	0.0005539

In this formulation, the number of species in each sample was expressed in terms of the expected number of species to normalize for the effects of salinity on the number of species present (Remane and Schlieper 1971). The number of species expected at a site was calculated from a third order polynomial fit to the 90th percentile of the distribution of the number of species vs. salinity for sites sampled during the 1990 Demonstration Project ($n = 34$, $r^2 = 0.96$, $p \leq 0.0001$) (see Section 4) . The polynomial was developed for a 3 ppt salinity running average of the number of species:

$$\text{Expected Number of Species} = 13.733 - (1.101 * \text{Salinity}) + (0.132 * \text{Salinity}^2) - (0.002 * \text{Salinity}^3)$$

The dimensionless scale of the discriminant score was converted to a range of 0 to 10 to clarify the presentation of results using the following algorithm:

$$\text{Benthic Index} = \frac{\text{Discriminant Score} - \text{Minimum Discriminant Score}}{\text{Discriminant Score Range}} \times 10$$

The threshold boundary between degraded and nondegraded benthos was set at 3.4 This was the midpoint between the mean of nondegraded and degraded discriminant values (see Section 5).

2.4.1.2 Fish

Visible pathological disorders were separated into two categories: those considered to be inducible by exposure to polluted conditions, and those that may not be the result of environmental conditions at the site, such as parasites. Only the first category, which included spots, ulcers, fin erosion, lumps or growths, skeletal malformations, opercular deformities, or abnormal condition of the eyes (e.g., cloudiness, hemorrhages, enlarged or sunken), was used in making status estimates.

The preliminary status evaluation contained two estimates of condition based on visible pathology: 1) the area with an abnormally high rate of pathological disorders, and 2) the prevalence of disorders per 1000 fish. The areal estimates were calculated using the same methodology as for the other parameters. The threshold criterion for an individual site was established at greater than 1% of the fish with visible pathological disorders. The threshold criterion also required observing disorders in at least two fish at the site to minimize false positives at sites where few fish were collected.

The estimate of prevalence of pathological disorders in fish differed from the generic methodology in that it involved a weighting factor in addition to the areal inclusion probability for each sampling site. This factor was relative abundance of fish caught at each site in the standardized trawl. This estimate involved two steps. First, a between-species prevalence rate for the station was calculated as the abundance-weighted average of the prevalence of each individual species. Next, the provincewide prevalence rate was estimated as a weighted average for the individual station values, where the weighting factor was the product of average catch-per-haul at the station and the station's inclusion probability. Prevalence rates for resource classes within the province were calculated in the same manner, except that inclusion probabilities for the class were used rather than for the province as a whole.

2.4.1.3 Sediment Toxicity

Estimates of area in the Virginian Province containing toxic sediments were based on the results of bioassays using the amphipod *Ampelisca abdita*; tests conducted with the freshwater amphipod *Hyalella azteca* were used only to evaluate the response of *Ampelisca* in low salinity habitats (see Section 5). Sediment toxicity was treated as a categorical variable in the status evaluation. A relative measure of toxicity was employed to facilitate comparisons between sites over a series of bioassays. Sediments were considered toxic if the survival of *Ampelisca* in test sediments was less than or equal to 80% of the survival observed in clean, control sediments, and if the percentage of survival in test and control sediments was significantly different ($p \leq 0.05$). These criteria were consistent with those established in USEPA/ACE (1991).

2.4.1.4 Sediment Contaminant Concentrations

Sediment contaminants were analyzed categorically, and concentrations higher than the ER-L values of Long and Morgan (1990) were classified as being of biological concern. Analyses were performed separately by contaminant and by classes of contaminants. The classes included contaminants having both natural and anthropogenic sources, such as metals; contaminants having mainly anthropogenic sources, such as polycyclic aromatic hydrocarbons (PAHs); and synthetic contaminants from industrial applications, such as polychlorinated biphenyls (PCBs). If any of the contaminants within a class were high, then the site was considered high for the whole class. Concentrations of all contaminants within the class had to be low to consider the site uncontaminated with that class of chemicals. The information from each site was used to estimate the status of contaminants by area throughout the province.

ER-L values represent concentrations at which some type of biological effects were noted in at least 10% of exposure studies reviewed by Long and Morgan (1990). These values were interpreted as concentrations at which biological effects begin to occur and were used to evaluate the extent of area where sublethal effects may occur. ER-L values are not available for all of the contaminants measured by EMAP, and data for chemicals without ER-L values were not included in the analysis. The estimates in Section 2, therefore, represent minimum estimates. Section 5 examines the effect of choosing alternative critical contaminant concentrations.

2.4.1.5 Sediment *Clostridium* Spores

The concentration of *Clostridium* spores in sediments was used to identify estuarine areas influenced by anthropogenic sewage inputs. *Clostridium* spores are ubiquitous in coastal marine environments and are found along the western Atlantic continental shelf at concentrations of about 10 colony forming units (CFU) per gram dry weight of sediment (Cabelli and Pedersen 1982; Duncanson et al. 1986); therefore, a threshold concentration had to be defined to separate background spore concentrations from elevated concentrations due to sewage inputs. For the assessment, *Clostridium* spore concentration was treated as a categorical variable, and sediments were considered to be influenced directly by sewage inputs if *Clostridium* concentrations were greater than 250 CFU per gram wet weight of sediment. This value is approximately equivalent to the background level of 100 CFU per gram dry weight of sediment identified in other studies (Cabelli, personal communication). Sediments with *Clostridium* spore concentrations less than or equal to 250 CFU per gram wet weight of sediment were not considered to be directly influenced by sewage inputs. These areas may, however, be indirectly influenced by farfield sewage inputs upstream or in other areas of the estuary from which spores were transported by currents.

2.4.1.6 Marine Debris

In the status evaluation presented in Section 2, marine debris was treated as a categorical variable, indicating the presence or absence of trash at a site. Only debris of anthropogenic origin (e.g. bottles, cans) was included in the estimates. Estimates were based on debris found either on the surface or on the bottom (i.e., collected in fish trawls).

2.4.1.7 Water Clarity

Water clarity was calculated using the vertical profiles of photosynthetically active radiation (PAR). Light extinction coefficients (k) were calculated from the profiles of PAR taken at each station. The light extinction coefficient was converted to Secchi disc depth (SDD) using the relationship k=1.7/SDD (Poole and Atkins 1929). A Secchi disc depth of 0.3 m (1 ft) was the threshold for water clarity.

2.4.1.8 Water Column Stratification

Density (sigma-t) was calculated using the standard equation of state for sea water (Millero and Poisson 1981), and the difference between surface and bottom water density was used as an indicator of water column stratification. A change in density greater than 2 was considered evidence for moderate to strong stratification, whereas a change less than 1 was considered evidence for weak stratification or no stratification.

2.4.1.9 Integration of Indicators

One objective of EMAP-Estuaries is to develop indices of environmental condition that integrate information from multiple indicators. Although individual response indicators provide information concerning specific aspects of environmental condition, overall statements regarding the condition of resources can be more useful to managers and nontechnical audiences. Single integrated statements can be communicated and understood more easily and are more appropriate for establishing and measuring progress towards environmental goals.

Analyses were completed to integrate information from indicators to focus on the two environmental attributes of interest (biological integrity and societal value). The integrated picture of biological integrity was developed by combining information from the benthic index and measures of fish pathology. Biological integrity was considered degraded at a site if the value of the benthic index was below the threshold value defining degraded benthic assemblages, or if the data for pathological disorders in fish exceeded the criteria of at least

two fish with pathological disorders and an incidence of 1% or greater. Biological integrity was considered nondegraded at a site if the benthic index indicated the presence of nondegraded benthic assemblages, and if the occurrence of pathological disorders did not exceed the criteria stated above.

Information about societal value was included in a similar manner. If trash was found at a site, *Clostridium perfringens* was present at greater than 250 CFU, or water clarity was less than 0.3 m Secchi disc depth, then societal value at the site was classified as degraded. If no trash was found, *Clostridium* was present only at background levels, and water clarity was greater than 0.3 m, then societal value was considered acceptable.

The indices of biological integrity and aesthetics were combined further to describe overall conditions within the estuaries of the Virginian Province. At each site, if either biological integrity or societal value was categorized as degraded, then the environmental conditions at the site were considered degraded. If both biological integrity and societal value were nondegraded, then environmental conditions at the site were considered nondegraded. This approach, in which each component indicator is weighted equally, is preliminary. EMAP presently is exploring ways to develop an overall index that weights indicators according to their relative importance to environmental managers and society.

2.4.2 Assessment Methods for Areal Estimation

Estimates of status for estuaries of the Virginian Province (Section 6) were made using data collected at the probability-based (base and supplemental) sampling sites. Each site was weighted according to the area associated with its sampling unit. For large estuaries, the area associated with each sample was equal to the size of the hexagon (280 km^2 for most sites; 70 km^2 in the Delaware Estuary, where supplemental sites were sampled). For large tidal rivers, the area was equal to the area of each river segment corresponding to the sampling site as measured using a geographic information system. For small estuarine systems, the area was equal to the area for that particular small system. For small estuarine systems having both base and supplemental sites (Back River, Elizabeth River, Indian River Bay, Mattopani River, Mullier River, and Mystic River), the area given to each was equal to the total area of the system divided by the number of base and supplemental sites in that system. To generate estimates within a resource class (large estuaries, large tidal rivers, small estuarine systems), the proportional weighting factors for each sample were calculated by dividing the individual area associated with each sampling site by the total area sampled in that resource class.

Large estuaries, large tidal rivers, and small estuarine systems were defined as sample strata in the sampling design; however, the stratification scheme was modified to define reporting classes because of perceived misclassification during stratification. Large tidal rivers originally

were designated on the basis of overall size and aspect ratio, but aspect ratio was considered only for rivers as a whole. The result was that river segments were ecologically dissimilar in some cases. In most cases, the biological resources in river segments were dependent upon conditions in upstream reaches of the river; however, in some cases, biological resources, especially benthic resources, were more dependent upon conditions downstream in the larger estuarine systems into which the river flowed.

These differences were particularly acute in the Potomac River, where the lower portion is more like an extension of the central axis of Chesapeake Bay than like a tidal river. The average width of the most upstream Potomac segment is only 1 km, whereas in the most downstream segment it averages more than 20 km. The channels in the most upstream segments have to be dredged to 8 m for navigation; natural channels exist to more than 20 m in the two lower segments. Water in the upper segments is not stratified and consists largely of runoff from upstream areas. The lower portion is highly stratified, with saltier, denser water from Chesapeake Bay flowing under the river runoff in classic, two-layer estuarine circulation (Lippson et al. 1979). Salinity differences from surface to bottom in the two lower Potomac segments were more than 5 ppt and 3 ppt during the 1990 sampling. No other tidal river segment had a difference of even 0.5 ppt.

Because of their size and ecological similarity to the adjacent Chesapeake Bay, the two lower stations in the Potomac River were included in the large estuary reporting class. This reclassification demonstrates the flexibility of the EMAP design. Because each sample has an associated area weight, relative inclusion probabilities can be calculated for almost any combination of sampling sites; thus, the large tidal river subpopulation need not be defined by the sampling stratum but can be defined on the basis of ecological properties. With this reclassification, the total area for the three classes was: large estuaries – 16,889 km^2, tidal rivers – 1809 km^2, and small estuarine systems – 4875 km^2.

To generate estimates across classes (strata), weights for sites within each class were adjusted so that the total of the weights for that class was equal to the total area represented by the sites within that class. On occasion, data were missing due to missed or lost samples, or failed quality control standards. In those circumstances, the total area represented by sampling did not equal the total area in the three resource classes, and adjustments were made for the evaluation of status to avoid undefinable sample area. A correction factor was applied to the previous weightings to compensate for missing data. The correction factor was the ratio of the area represented by all possible samples to the area represented by all samples for which data were available. Correction factors were calculated for each reporting class. For the small estuarine systems class, the area representing all possible samples was the area of all estuaries in the list frame of small systems. The 1990 Demonstration Project sampled approximately one-fourth (32 out of 137) of these estuaries.

Estimates in the preliminary evaluation were based only on data collected during sampling intervals 2 (19 July to 31 August) and 3 (1 September to 23 September). The first occurrence of data from each site was selected for inclusion in the estimates. Data from sampling interval 1 were excluded from the calculations because examination of the data revealed that exposure conditions were not as severe then as they were later in the summer (see Section 6-1).

2.4.3 Procedures for Estimating Precision

The approximate 90% confidence intervals for the province were calculated based on the assumption that the CDF estimates were distributed normally. The confidence intervals were obtained by adding and subtracting 1.645 times the estimated standard error (square root of the variance) to the estimated CDF value.

For small estuarine systems, estimates of CDFs and associated variances were computed based on a random selection of small systems within the province, with replicate samples taken from a subset of the selected systems (Cochran 1977). The resulting CDF estimate is:

$$\hat{P}_{S,x} = \frac{\sum_{i=1}^{n} A_i \bar{y}_i}{\sum_{i=1}^{n} A_i}$$

where

$\hat{P}_{S,x}$ = CDF estimate for value x

$\bar{y}_i$ = $\frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$

m_i = number of samples at small system i

y_{ij} = $\begin{cases} 1 \text{ if response is less than x} \\ 0 \text{ otherwise} \end{cases}$

A_i = area of small system i

n = number of small systems sampled

Since replicate samples were obtained only at a subset of the sampled small estuarine systems, the formula for the estimated variance taken from Cochran (1977 eq. 11.30) was modified to produce the following estimate of the approximate mean squared error (MSE) of the CDF estimate:

$$MSE(\hat{P}_{s,x}) = \frac{\frac{N^2}{n}(1-f_1)\frac{\sum_{i=1}^{n} A_i^2(\bar{y}_i-\hat{P}_{s,x})^2}{n-1} + \frac{N}{n*}\sum_{i=1}^{n*}\frac{A_i^2 S_{2i}^2}{m_i}}{A^2}$$

where

N	=	number of small systems in the province (137)
f_i	=	n/N
n^*	=	number of small systems with replicate samples
S_{2i}^2	=	$\frac{\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_i)^2}{m_i-1}$
A	=	the total area of small systems in the province (4875 km^2)

Estimates of CDFs for large tidal rivers were obtained by applying Horvitz-Thompson estimation (Cochran 1977) with selection probabilities being inversely related to station area. Data from all base stations and from supplemental stations sampled in the Delaware River were used in this analysis. Areas of base stations in the Delaware River were adjusted to reflect the inclusion of the supplemental stations. Estimates of CDFs were:

$$\hat{P}_{T,x} = \frac{1}{A}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$$

where

$\hat{P}_{T,x}$	=	Estimated CDF at value x
y_i	=	$\begin{cases} 1 \text{ if response is less than x} \\ 0 \text{ otherwise} \end{cases}$
π_i	=	inclusion probability for station i (1/area)
A	=	total area of the sampled tidal rivers
n	=	number of stations sampled

Variance estimators that are commonly used with general probability samples, such as the Horvitz-Thompson or Yates-Grundy (Cochran 1977), require that all joint event probabilities π_{ij} must be non-zero, which is not the case for systematic random designs, such as EMAP's.

Estimates derived from such designs do not have unbiased estimators of variance. Instead, the variance is approximated based on a model or set of assumptions. The variance approximation to estimate confidence intervals presented in Section 6 for the status estimates was derived from the Yates-Grundy estimator:

$$\hat{var}(\hat{p}_{T,x}) = \frac{1}{A^2} \sum_{i=1}^{n} \sum_{j>i}^{n} \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

where

π_{ij} = probability that sites i and j are selected for sampling

and

$$\pi_{ij} = \frac{2(n-1)\pi_i \pi_j}{2n - \pi_i - \pi_j}$$

The resulting approximation no longer depends on the joint event probabilities, but only on the first order inclusion probabilities.

The approximation was developed by Stehman and Overton (1989) and was derived under the assumption that the population is randomized between successive draws. The use of this approximation based on the Yates-Grundy variance estimator amounts to assuming that the grid-based sample is nearly a simple random sample. Simulation studies (Overton and Stehman 1987; Stehman and Overton 1987a,b) have demonstrated that the approximation performs well in EMAP-like sampling circumstances. The systematic sample used in large systems and tidal rivers should provide more precise estimates than simple random sampling, so that the approximation will provide conservative estimates of precision.

Estimates of CDFs for large systems also were obtained by applying Horvitz-Thompson estimation with selection probabilities being inversely related to station area. Data from all base stations in the Virginian Province and from supplemental stations in the Delaware Bay were used in this analysis. Areas of all stations in the Delaware Bay were assumed to be 70 km^2 due to the inclusion of the supplemental stations. Areas for other large estuary base stations were assumed to be 280 km^2. The areas for the two lower Potomac River stations included in the large estuary class were the areas of the corresponding tidal river segments. Formulae for the CDF estimates and corresponding variances are analogous to those presented for large tidal rivers.

Estimates of CDFs for a particular geographic system within the province (e.g. Chesapeake Bay system) were obtained by applying the above procedures to the small estuarine systems, tidal rivers, and large estuaries sampled within that geographic system. Estimates of the CDFs for the entire province or for a geographic system within the province were computed as weighted averages of the relevant station class CDFs:

$$\hat{P}_x = W_S\hat{P}_{S,x} + W_T\hat{P}_{T,x} + W_L\hat{P}_{L,x}$$

where

W_S, W_T, W_L = relative areas for small systems, tidal rivers, and large estuaries, respectively.

The variance of the estimate is:

$$\hat{var}\,(\hat{P}_x) = W_S^2 var(\hat{P}_{S,x}) + W_t^2 var(\hat{P}_{T,x}) + W_L^2 var(\hat{P}_{L,x})$$

In applying these procedures, variance estimation was based on the assumption of a fixed sample size within each resource class. For large tidal rivers and large estuaries, the sample size is a random element depending on the position of the sampling grid. This variance component has not been incorporated into the estimation of variances of CDFs.

Estimating the percentage of fish with pathological disorders represents a different analytical problem because it does not involve an estimate of area. The calculation was treated as two-phase sampling. For each sample station, the calculated estimate of the proportion of fish with pathologies was:

$$\hat{P}_i = \sum_{s-1}^{m} W_{is}\hat{P}_{is}$$

where

$\hat{P}_i$ = the estimated proportion of fish pathology at station i

m = number of species caught at station i

W_{is} = n_{is}/n_i

n_{is} = number of fish of species s caught at station i

n_i = total number of fish caught at station i

$\hat{P}_{is}$ = x_{is}/e_{is}

e_{is} = number of fish of species *s* that were examined for pathologies at station *i*

x_{is} = number of examined fish of species *s* that had pathologies at station *i*

Applying the estimation methods of two-phase sampling, the estimated variance of this estimate was calculated by (Cochran 1977, eq. 12.24):

$$\hat{VAR}(\hat{P}_i)=\sum_{s=1}^{m}\frac{W_s^2\hat{P}_i(1-\hat{P}_i)}{(e_{is}-1)}+\frac{\sum_{s=1}^{m}W_s(\hat{P}_{is}-\hat{P}_i)}{n_i}$$

For large estuaries, the estimate of the fish pathology rate was calculated by:

$$\hat{PL}=\frac{\sum_{i=1}^{n_L}\frac{1}{\pi_i}\bar{n}_i\hat{P}_i}{\sum_{i=1}^{n_L}\frac{1}{\pi_i}\bar{n}_i}=\frac{A}{B}$$

where

n_L = number of large estuaries sampled

$\bar{n}_i$ = average catch per unit effort at station *i*

The estimated variance of the estimate was calculated by:

$$\hat{VAR}(\hat{PL})=\hat{VAR}(\frac{A}{B})=\frac{A^2}{B^2}(\frac{\hat{VAR}(A)}{A^2}+\frac{\hat{VAR}(B)}{B^2}-\frac{2\hat{COV}(A,B)}{AB})$$

where

$$\hat{VAR}(A)=\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{\bar{n}_i\hat{P}_i}{\pi_i}-\frac{\bar{n}_j\hat{P}_i}{\pi_j}\right)^2+\sum_{i=1}^{n}\frac{\bar{n}_i^2 VAR(\hat{P}_i)}{\pi_i^2}$$

$$\hat{VAR}(B)=\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{\bar{n}_i}{\pi_i}-\frac{\bar{n}_j}{\pi_j}\right)^2$$

$$\hat{COV}(A,B)=\sum_{i=1}^{n}\frac{(1-\pi_i)}{\pi_i^2}\bar{n}_i^2\hat{P}_i+\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}}\right)\bar{n}_i\bar{n}_j\hat{P}_i$$

The estimate of the proportion of fish pathology for large tidal rivers ($\hat{PT}$) and associated variance ($\hat{VAR}(\hat{PT})$) was calculated analogously to the estimation for large estuaries.

For small systems, the estimated proportion of fish pathology was estimated by:

$$\hat{PS}=\frac{\sum_{i=1}^{n_s}R_i\hat{P}_i}{R}$$

where

$\hat{PS}$	=	the estimate of fish pathology for small estuarine systems
n_s	=	the number of small estuarine systems sampled
R_i	=	$A_i\bar{n}_i$
A_i	=	area of system i
R	=	$\sum_{i=1}^{n_s}R_i$

The estimated variance of the estimate (Cochran 1977, eq. 11.30) was calculated by:

$$\hat{VAR}(\hat{PS})=\frac{n}{R^2}(1-f)\frac{\sum_{i=1}^{n_s}R_i^2(\hat{P}_i-\hat{PS})^2}{n-1}+\frac{n}{NR^2}\sum_{i=1}^{n_s}R_i^2 VAR(\hat{P}_i)$$

Provincewide estimates for proportion of fish pathology were calculated by weighted averages of small estuarine systems, large tidal rivers, and large estuaries estimates.

SECTION 3
EVALUATION OF LOGISTICAL FEASIBILITY AND QUALITY ASSURANCE

One of the primary objectives of the 1990 Demonstration Project was to evaluate the feasibility of collecting, within a limited sampling period, the kinds and volume of data required to produce a regional evaluation of the status of estuarine ecosystems in the Virginian Province. This section uses the results of the 1990 Demonstration Project to address two fundamental questions pertaining to logistics and quality assurance for future EMAP monitoring in estuaries:

- Could the data required for developing regional status estimates be collected with the level of effort and sampling methods employed in the Demonstration Project?

- Does the sampling plan ensure collection of data that satisfy criteria defining suitability for the intended use (i.e., quality assurance)?

The section also identifies major logistical impediments encountered during the 1990 Demonstration Project and provides recommendations for improvement in future years. Logistical successes and failures are identified from both the sample collection and processing perspectives for each indicator.

Only results from the second and third sampling intervals (19 July to 23 September) were considered in evaluating the logistical feasibility of the EMAP sampling plan. Together, these two intervals represent a "typical" sampling effort for future years of EMAP monitoring in the Virginian Province in terms of duration, the skill and experience of the sampling crews, and the types and number of expected station visits (including an allowance for station revisits, if necessary). Although it produced some usable data, the first sampling interval during the 1990 Demonstration Project was used primarily for additional crew training and refinement of procedures. As a result, only about half of the site visits originally scheduled for interval 1 were completed successfully. Including this first sampling interval in the evaluation of logistical feasibility would needlessly underestimate the effectiveness of the logistical plan for use in future implementation of the program, when sampling procedures will be fully tested and refined, and crews will be fully equipped and experienced.

In general, the field crews succeeded in obtaining the required data within the allotted time. The overall effectiveness of the 1990 sampling plan is reflected in the high percentage of stations for which usable data were obtained for all parameters (Table 3-1). The number of stations with usable data is sufficient for addressing the major objectives of the 1990 Demonstration Project, including the evaluations of sampling design, indicator

Table 3-1. Status of sample collection during the 1990 Virginian Province Demonstration Project

Parameter	Expected Number of Stations Sampled	Percent of Stations[a]			
		Collected	Received[b]	Analyzed	Passed QA
Water Quality:					
CTD Profiles	217	99	96	96	86
Suspended Solids	23	100	100	100	100
Chlorophyll *a*	217	95	95	0	0
Water Column Toxicity	23	100	100	100	100
Clostridium	160	95	95	93	93
Datasonde Deployments	30	93	90	90	73
Sediment:					
Inorganic Chemistry	160	95	95	93	93[c]
Organic Chemistry	160	95	95	93	73[c]
Grain Size	160	96	94	93	93
Sediment Toxicity	160	96	95	91	88
Benthic:					
Infaunal Assemblages	217	96	96	96	96
Silt-clay Content	217	96	95	95	95
Sediment Profile Images	20	85	85	85	85
Fish:					
Assemblages/Pathology	160	89	89	89	89
Chemistry	160	70	69[d]		
Bivalve:					
Chemistry	126	29	29	10	10

[a] Values represent the percent of stations sampled in intervals 2 or 3 for which the various steps were completed successfully.
[b] This accounts for samples lost at collection or during shipment.
[c] Based on expected number of analyte results rather than expected number of stations.
[d] Tissue samples exceeded holding times and were not processed.

suitability, and estuarine status. Although data collection rates were uniformly high, there were some notable QA/QC deficiencies in the field sampling component of the Demonstration Project. The following sections offer detailed descriptions of these deficiencies and provide recommendations for overcoming them in future years.

It is noteworthy that in over 5,000 man-hours of field activity in three months, there were no serious injuries or significant losses of equipment. Overall, these results confirm that small boats (24 ft) can be used to collect samples in estuaries safely and effectively. It must be stressed that there was considerable emphasis on safety procedures during crew training (including first aid and CPR) and throughout field operations; furthermore, there was an appropriate allowance in the schedule for inclement weather and equipment malfunction, and back-up equipment was available at all times throughout the 1990 Demonstration Project.

3.1 WATER QUALITY PARAMETERS

Both the vertical profile and continuous near-bottom water quality measurements were recorded internally, using the Seabird CTD and Hydrolab DataSonde 3, respectively; the resultant electronic files were downloaded, stored, and transferred using the field computers. A major logistical objective was to determine the feasibility of deploying and retrieving moored Hydrolabs on a 10-day cycle, with the ultimate goal of obtaining continuous 70-day records at 30 sites throughout the province.

The field crews successfully retrieved approximately 90% (110 of 122) of the Hydrolabs deployed throughout the 1990 Demonstration Project. Despite deployment of units at a wide variety of locations, only eight were lost permanently. Several others were displaced from their mooring sites but were found later or retrieved by divers. Units that were lost permanently were generally from a few high-traffic sites.

The high success rate in retrieving deployed Hydrolabs and the general reliability and ease of operation of the instruments indicate the overall feasibility of using Hydrolabs to obtain continuous near-bottom data in future years. Some Hydrolab records were lost, however, as a result of internal electronic problems or failure to communicate with the field computers. Field QC checks and subsequent review of the continuous records suggested that biological fouling of the dissolved oxygen sensors, which typically occurred several days after deployment, frequently diminished the accuracy of readings. This suggests the need to shorten the deployment interval in future years.

CTD files were received successfully from more than 90% of stations (Table 3-1). At less than 10% of stations, CTD files were not available for final QA/QC review due either to failure to deploy the instrument, or inability to download data because of intermittent electronic or field computer software problems. Diskettes containing CTD files from a small percentage of

stations were lost during shipment. These data could not be recovered because the files were deleted from the field computers after the diskettes were shipped. A greater level of effort devoted to troubleshooting and debugging the computer software and better crew training prior to field operations should result in lower rates of data loss in future years.

The original plan for QA/QC review of the CTD profile data was to compare daily Winkler DO measurements against the simultaneous CTD reading as a check on the accuracy of the latter. This check was unsuitable for several reasons. First, there is considerable doubt about the accuracy of the Winkler titrations performed by the field crews due to insufficient training in this technique. Second, there is evidence that exposure to excessive heat caused the Winkler chemicals to become unstable as the summer progressed. Third, and most important, the Winkler accuracy check was performed only once a day while the CTD was at the surface; therefore, its ability to detect inaccurate readings during deployment of the instrument on station was limited.

The overall accuracy of each of the Seabird CTDs was confirmed in a laboratory calibration check immediately following the completion of the 1990 sampling effort. In the absence of a reliable, daily, field QC check, the vertical plot of each parameter in the CTD data files (e.g., temperature, salinity, pH, dissolved oxygen, PAR, percent transmission, and chlorophyll *a* fluorescence) was reviewed visually for consistency between descending and ascending profiles. Rejection of the water column profile data from 20 stations (Table 3-1) is attributed to problems with physical deployment of the CTDs at specific stations. At most of these stations, sharp, anomalous drops in both dissolved oxygen and salinity, and lack of agreement between the downcast and upcast profiles suggested that sediments drawn into the CTD pump system upon contact with the bottom may have clogged it completely. Anomalous dissolved oxygen and salinity readings helped to identify this problem because only these two sensors require flowing water to obtain proper readings. The manual review of CTD data resulted in rejection of profiles that 1) indicated that field crews did not allow enough time for the dissolved oxygen sensor to equilibrate at depth before taking measurements, or 2) exhibited improbable (e.g., supersaturated) or unstable readings as a result of intermittent electronic problems.

Several recommendations could improve the success rate for obtaining usable vertical profiles with the Seabird CTD in the future. First, to prevent clogging problems, field crews should avoid letting the instrument contact the bottom during deployment. Second, a field QC check should be done at every station and should provide an independent measure of bottom dissolved oxygen concentration for comparison with the CTD reading. A Nisken bottle could be used to obtain a bottom water sample from which to measure dissolved oxygen concentration using either Winkler titration or a second instrument (e.g., a hand-held YSI meter). This check would reveal inaccurate readings associated with improper deployment of the CTD and provide instantaneous confirmation of acceptable vertical profiles. Third, field crews should undergo more intensive training and be given greater responsibility for determining the acceptability of CTD profiles on station. This would provide greater

opportunity for the crews to redeploy the instrument if an initial cast is suspected to be unacceptable. Finally, back-up instruments should always be available to replace field instruments in need of service or recalibration.

Usable data on the concentration of suspended solids and water toxicity were obtained at most stations despite the stringent preservation requirements (shipment below 4°C) and narrow holding-time limits (less than 48 hours after sample collection for water toxicity). In contrast, none of the chlorophyll *a* samples were suitable for processing. The filters that arrived at the laboratory were either thawed or too wet due to inadequate training of field crews in use of the simple syringe filtering apparatus. Measurement of chlorophyll *a* concentration will be reassessed for its suitability as an indicator for estuaries.

3.2 SEDIMENT QUALITY AND BENTHIC COMMUNITY PARAMETERS

Sediment samples for sediment quality and benthic community analyses were obtained successfully at over 95% of the scheduled stations (Table 3-1). The roughly 5% of sites where samples could not be obtained were characterized by rocks or cobbles; the sampling gear used in the Demonstration Project is not designed to work in these bottom types. Chemistry, grain size, and toxicity samples from an additional 2% to 5% of the stations either were lost in shipping or could not be processed upon receipt at the laboratory due to container breakage. Sturdier containers will be used in the future, particularly for the sediment toxicity and chemistry samples; furthermore, greater emphasis will be placed on instructing field crews in packing techniques to avoid breakage.

Sediment profile images were collected from 20 stations in the northern half of the Virginian Province. Organism-Sediment Index (OSI) values could not be calculated from data collected at three of the stations (69, 77, 93) because of poor camera penetration. At Station 69, the sand content was greater than 75%, and the surface dwelling gastropod *Crepidula* was one of the dominant forms. No grain size data are available for Station 77; sand content at Station 93 was greater than 95%. OSI data were collected from the remaining four high sand content (greater than 75%) stations in this survey. It appears from these data that sediment profile imaging may be restricted at sites that have very high sand content or heterogeneous, unconsolidated deposits.

Average laboratory error rates of less than 4% were achieved for sediment grain size and benthic analyses (including sorting, species identification and enumeration, and biomass). These error rates represent the cumulative results of a series of internal QA/QC checks designed to ensure consistent production of usable data. Error rates above acceptable limits triggered a series of corrective actions, including re-analysis of samples. Because the overall laboratory processing error rate was low for the sediment grain size and benthic analyses, the

final data were deemed usable for a high percentage of the stations from which samples were collected.

Acceptability of the sediment chemistry data depended upon the ability of the laboratory to perform the analyses within pre-established control limits. For the inorganic analyses, the laboratory generally was able to meet control limits for the required QA/QC samples (e.g., calibration check samples, blanks, matrix spikes). Laboratory results for Standard Reference Materials (SRMs), which were analyzed along with every sample batch as a QC check on both accuracy and precision, are presented in Table 3-2. These results indicated acceptable recoveries for the metals analyzed by inductively coupled plasma-atomic emission spectrometry (ICP-AES; Al, Cr, Cu, Fe, Mn, Ni, Pb, Zn) and acceptable, although somewhat lower and more variable, recoveries for the metals analyzed by graphite furnace atomic adsorption (GFAA; As, Cd, Sb, Se, Sn). The increased variability in GFAA analyses was due both to the lower concentrations of these metals in the SRM and the less rigorous digestion procedure used; therefore, some of the metal concentrations determined by GFAA may slightly under- or overestimate the true amount present in the sample. Silver was not detected in most of the samples; however, the laboratory's detection limit of 1 ppm was above the EMAP-E target detection limit of 0.1 ppm for this metal.

The expected number of analyte results for inorganics was 2235, based on analysis of 149 samples for 15 different metals. Because results for QA/QC samples generally fell within the control limits established for the program, a high percentage (93%) of the associated metals data was deemed to have "passed QA" (Table 3-1) and was used to evaluate sediment contamination in the Virginian Province. Only results for silver failed to meet QA criteria, due to the laboratory's inability to achieve the required detection limit for this metal.

The laboratory that analyzed the 1990 sediment samples experienced some early difficulties with the organic analyses, however, the final QA/QC results for the SRM (Table 3-3) suggested that the analyses were performed within acceptable control limits for accuracy and precision. Very high and variable rates of recovery were experienced for the pesticides heptachlor epoxide, cis-chlordane, trans-nonachlor, and 4,4'-DDT, which were present in the SRM at concentrations very close to the laboratory's detection limit. Such a high degree of uncertainty and variability is to be expected from any laboratory trying to quantify organic analytes at concentrations close to the detection limit in a complex matrix like SRM 1941 (i.e., Baltimore Harbor sediment). As analyte concentrations increase above the detection limit, there is a concomitant increase in a laboratory's ability to detect and quantify the compounds accurately (Keith 1991). This is reflected in the SRM 1941 results for 1990: good recoveries were achieved for PAHs and PCB congeners with high SRM concentrations relative to the laboratory's detection limit. Based on these results, it is reasonable to conclude that the 1990 organics data is reliable for samples having similarly elevated organic contaminant concentrations.

Table 3-2. Results for SRM 2704 (Buffalo River Sediment) used as a set control for the 1990 EMAP-E sediment inorganic analyses.

Element	Average[a]	Stdv[b]	CV[c]
ICP-AES Metals (n = 18 analysis sets or "batches")			
Al	96	1.8	1.9
Cr	87	2.7	3.1
Cu	96	2.1	2.2
Fe	88	1.6	1.8
Mn	94	2.1	2.2
Ni	90	5.5	6.2
Pb	92	4.3	4.6
Zn	96	1.5	1.6
GFAA Metals (n = 18 analysis sets)			
As	78	4.1	5.3
Cd	100	6.8	6.8
Sb	79	11.4	14.4
Se	97	12.1	12.5
Sn	80	29.4	36.7

(a) Average percent recovery relative to the SRM certified value
(b) Standard deviation of the percent recovery values
(c) Coefficient of variation of the percent recovery values

The laboratory's failure to meet the target detection limits consistently is a major deficiency in the organics data sets (Table 3-4). The analytical method resulted in high detection limits, and detection limits varied because the laboratory analyzed a different amount (i.e., dry weight) of sediment from each sample. As a result, the target analytes were not detected in a large number of samples, and the "calculated" detection limits (i.e., the theoretical concentration of each analyte necessary for detection) differed significantly from sample to sample (Table 3-4).

Table 3-3. Results for SRM 1941 (Baltimore Harbor sediments) used as a set control for the 1990 EMAP-E sediment organic analyses.

Compound[a]	Average[b]	Stdv[c]	CV[d]
PAHs (n = 19 analysis sets or "batches")			
Phenanthrene	95.8	20.6	21.5
Anthracene	69.1	17.1	24.7
Fluoranthene	97.7	22.8	23.3
Pyrene	85.0	18.0	21.2
Benz[a]anthracene	93.8	21.6	23.0
Benzo[b+k]fluoranthene	101.9	17.6	17.3
Benzo[a]pyrene	63.1	16.3	25.8
Perylene	62.3	16.8	27.0
Benzo[ghi]perylene	84.9	23.3	27.4
Indeno[1,2,3-cd]pyrene	120.8	29.5	24.4
PCBs/pesticides (n = 15 analysis sets)[e]			
PCB 18	80.6	14.1	17.5
PCB 28	53.5	11.3	21.1
PCB 52	96.2	23.4	24.3
PCB 66	66.2	11.6	17.5
PCB 101	69.2	15.4	22.2
PCB 118	99.3	15.4	15.5
PCB 153	95.2	15.3	16.1
PCB 105	91.0	18.9	20.8
PCB 138	72.1	15.4	21.3
PCB 187	78.5	17.7	22.5
PCB 180	89.5	17.6	19.7
PCB 170	81.6	22.8	27.9
PCB 195*	135.2	39.1	28.9
PCB 206*	97.2	29.7	30.5
PCB 209	88.1	19.9	22.5
Heptachlor epoxide*	238.0	78.3	32.9
Alpha-Chlordane*	304.0	55.3	18.1
Trans-Nonachlor*	547.0	835.0	152.0
p,p'-DDE	97.3	31.3	32.2
p,p'-DDD	86.2	23.2	26.9
p,p'-DDT*	202.0	141.0	69.8

[a] SRM 1941 has certified concentrations for only a subset of the PAH compounds of interest to EMAP-E.
[b] Average percent recovery relative to the SRM certified value
[c] Standard deviation of the percent recovery values
[d] Coefficient of variation of the percent recovery values
[e] SRM 1941 only lists "non-certified" or informational values for this group of PCB congeners and pesticides (* = concentration in the SRM is < 10 times the target detection limit).

Table 3-4. Range in detection limits (in ng/g dry weight) reported for organic compounds in 1990 sediment samples. The target detection limits were 10 ng/g for each PAH compound and 0.5 ng/g for each PCB congener and pesticide.

	Minimum	Maximum	Median
Polycyclic Aromatic Hydrocarbons (PAHs)			
Acenaphthene	21	207	34
Anthracene	17	121	28
Benz(a)anthracene	17	72	28
Benzo(a)pyrene	23	151	38
Benzo(e)pyrene	23	153	37
Biphenyl	23	150	36
Chrysene	22	72	35
Dibenz(a,h)anthracene	24	252	43
2,6-dimethylnaphthalene	24	156	38
Fluoranthene	16	114	24
Fluorene	25	176	43
2-methylnaphthalene	25	162	39
1-methylnaphthalene	23	150	34
1-methylphenanthrene	13	86	21
Naphthalene	30	54	39
Perylene	27	189	46
Phenanthrene	16	44	26
Pyrene	15	39	22
Benzo(b+k)fluoranthene	22	145	33
Acenaphthlylene	22	212	38
Benzo(g,h,i)perylene	31	325	55
Ideno(1,2,3-c,d)pyrene	26	249	43
2,3,5-trimethylnaphthalene	23	219	38
DDT and its metabolites			
o,p'-DDD	0.13	1.93	0.24
p,p'-DDD	0.12	6.10	0.20
o,p;-DDE	0.10	1.11	0.18
p,p'-DDE	0.04	0.45	0.07
o,p'-DDT	0.12	1.26	0.22
p,p'-DDT	0.18	3.22	0.58

Table 3-4. Continued			
	Minimum	**Maximum**	**Median**
Chlorinated pesticides other than DDT			
Aldrin	0.10	1.78	0.27
Alpha-Chlordane	0.09	1.16	0.19
Trans-Nonachlor	0.04	0.87	0.07
Dieldrin	0.04	0.52	0.08
Heptachlor	0.10	1.47	0.19
Heptachlor epoxide	0.08	1.85	0.19
Hexachlorobenzene	0.03	7.23	0.09
Lindane (gamma-BHC)	0.16	27.5	0.64
Mirex	0.03	1.93	0.08
18 PCB Congeners			
PCB 08	0.08	4.46	0.63
PCB 18	0.37	5.89	0.94
PCB 28	0.08	1.03	0.17
PCB 44	0.06	1.50	0.17
PCB 52	0.11	2.70	0.38
PCB 66	0.09	1.01	0.18
PCB 101	0.12	1.39	0.20
PCB 105	0.07	0.60	0.14
PCB 118	0.06	0.65	0.12
PCB 128	0.12	1.62	0.23
PCB 138	0.11	1.31	0.18
PCB 153	0.11	1.03	0.19
PCB 170	0.09	2.15	0.32
PCB 180	0.11	1.30	0.19
PCB 187	0.08	0.72	0.13
PCB 195	0.10	1.23	0.19
PCB 206	0.10	1.38	0.20
PCB 209	0.12	1.09	0.20

The expected number of results for organics was 8344, based on analysis of 149 samples for 23 PAH compounds, 15 pesticides, and 18 PCB congeners. Any results reported as "not detected" in samples for which the laboratory failed to meet the EMAP-E target detection limits (i.e., 10 ppb for individual PAHs, 0.5 ppb for individual pesticides/PCBs) were considered to have "failed" QA; therefore, the overall success rate for the organics analyses was 73% (Table 3-1). If the target detection limits had been achieved and consistent sample sizes had been used, quantifiable amounts of the organic analytes of interest probably would have been

detected in most of the 1990 samples, and a higher percentage of results would have passed QA.

Concentrations near the program detection limits represent the "background" levels of many of the EMAP-E organic contaminants in the estuarine sediments of the Virginian Province. The difficulty encountered in consistently meeting the target detection limits impairs our ability to make quantitative evaluations of the percent of bottom area in the Virginian Province that is contaminated with organic compounds. Chemicals may have been present that were not detected and quantified. This limits the comparability of the 1990 chemistry data with other data sets for which lower detection limits were achieved.

A number of corrective actions and changes in methodology have been implemented by the laboratory responsible for the 1990 analyses. For example, future analyses will be performed using a constant dry weight of sample, which should result in a consistent detection limit. The 1990 experience helps to illustrate some of the difficulties inherent in performing low-level (i.e., low parts-per-billion) analyses of organic contaminants in complex samples from estuarine environments.

Of the 160 sediment toxicity samples scheduled, 146 were processed. Results from 35 of these samples are qualified (i.e., "flagged") in the database because average survival of the test organism (the amphipod *Ampelisca abdita*) in the controls for these stations was less than the pre-established limit of 90%. In 30 of these qualified results, however, control survival was greater than 86%, and they were used in the preliminary evaluation. When the control survival requirement was reduced to 85%, only five stations (i.e., less than 4% of the stations tested) were unusable due to unacceptable control survival. The original 90% survival criterion, which was adopted from the ASTM (1991) procedure for the amphipod *Rhepoxynius abronius*, may not be achievable for all species, especially those for which a large database is being formed only now. For example, the ASTM control survival requirement for *Hyalella* is 80%. Data loss can be avoided to an even greater extent in the future by collecting enough sediment to allow for retesting in the event of unacceptable control survival; however, even without this additional sample effort, the 1990 Demonstration Project results indicate that the *Ampelisca* test can provide usable sediment toxicity data for a high percentage of stations.

Clostridum spore concentrations in sediments were determined from subsamples taken from the same homogenized sediment samples used for the determination of sediment contaminant concentrations and sediment toxicity. Subsamples were collected when sediment samples were thawed for extraction of contaminants. Established methods for *Clostridium* spore counts require conducting assays on the day samples are thawed; however, because of the number of sediment samples scheduled for contaminant extraction each day, not all subsamples for *Clostridum* counts could be processed on the day they were thawed. Recognizing this problem, laboratory personnel obtained some *Clostridium* samples by removing chips of sediment from samples prior to thawing and keeping them frozen until they could be processed. The time delay prior to processing some thawed *Clostridium* samples and analysis of subsamples that had not been homogenized with the whole sample after

thawing (i.e., frozen chips) both represent departures from established methods. Samples so processed were flagged with appropriate quality assurance codes in the database.

Additional analyses were conducted to determine how these departures from established procedures influenced spore counts. Tests demonstrated that storing thawed samples for up to four days did not significantly reduce *Clostridium* spore concentrations. Since no thawed samples were stored for more than four days, the time delay probably did not influence spore counts. The additional analyses demonstrated that the frozen sediment chips contained about 43% fewer spores than subsamples drawn from sediments that were homogenized after thawing. Spore concentrations were determined from frozen sediment chips for 25 of 177 samples. Doubling spore concentrations to correct for using frozen sediment chips would have changed the interpretation for only one of those 25 samples; consequently, no corrections were made in the data set, and all samples were used in the evaluation presented in Section 2.

3.3 FISH AND BIVALVE SAMPLING

A successful trawl was defined as one in which the crew was able to deploy and "fish" the 16-m (footrope length) high-rise trawl net for 10 ± 2 minutes, regardless of whether fish were caught. This condition was met at 143 (89%) of the 160 stations where fish trawls were scheduled (Table 3-1). Fish were caught at 140 (98%) of the 143 stations where trawls were successful. These high percentages indicate the logistical feasibility of using the chosen equipment and standard trawl duration to catch fish at a variety of estuarine locations.

Despite the high rate of fish sampling success, the Demonstration Project identified some limitations of the chosen methodology. Trawls were not even attempted at 15 stations. This includes 11 sites that were too deep (greater than 80 ft) for proper deployment of the gear, 2 shallow sites that lacked sufficient room to maneuver the boat, and 2 sites where equipment failure prevented deployment of the net. At the remaining two stations where trawls were unsuccessful, trawling was attempted but had to be aborted because the net either filled with weeds or got entangled on an underwater obstruction.

Target fish species were collected and saved for chemical analyses at 112 (78%) of the 143 stations where trawls were successful; this constitutes 70% of the stations where trawls were scheduled (Table 3-1). At 13 of the sites where no target fish were caught, only one trawl was performed. The success rate for capturing target fish could improve in the future if additional sampling effort is expended at such sites. This might involve doing an additional trawl right away or returning to the site on another day. At an additional 13 sites, the field crews caught only scup or butterfish. These species initially were included among the target fish but were eliminated later to reduce handling time on the boat. They have since been reinstated as target species, which should result in a higher overall capture rate for target species in the future.

Some of the target fish samples, representing collections from ten stations, were lost because they were not saved or were not shipped to the laboratories. The entire collection of target species was lost at only two of these stations. Improvements in the sample tracking system and in crew training have been developed to avoid this loss in the future. The 110 stations (69% of the anticipated stations; Table 3-1) for which chemistry samples were received by the analytical laboratory would have been sufficient for assessing the feasibility of the fish contaminant indicator; however, the fish tissue samples were scheduled for analysis following the completion of the sediment samples, and delays in finalizing the sediment organic analyses resulted in exceeding the recommended holding time for the fish tissue. For this reason, fish contaminant analyses were not performed on the 1990 samples.

Besides saving fish for chemical analyses, the field crews recorded species composition and abundance for each trawl, measured the length of each individual, and examined each fish for visible external pathologies. Approximately 150 specimens, representing 48 species, were sent to the laboratory for independent verification of the field identification. The laboratory taxonomic expert had to identify 11 of the 150 specimens because they were not identified by the field crews. Nine specimens were identified incorrectly by the field crews; thus, 130 of the 150 (87%) individuals received for taxonomic verification were identified correctly by the field crews.

Some of the fish that either were not identified or were misidentified in the field were uncommon or exotic species (e.g., blue runner, striped cusk eel, spotted eagle ray) of which only a few representatives were collected throughout the summer. The crews consistently identified the most abundant species correctly, including the target species. The only exceptions to this were identifications of catfish (Family Ictaluridae) and clupeids (Family Clupeidae, including menhaden, herring and anchovies). These misidentifications were infrequent and limited to within-family errors (e.g., blue catfish identified as white catfish, alewife identified as Atlantic herring). These results will be used to refine and focus crew training in future years.

At 30 of the 140 stations where fish were collected, at least one fish identified by the field crews as having one or more visible external pathologies was sent to the EPA's Gulf Breeze laboratory for confirmation of the pathologies. A total of 572 fish identified as having no visible external pathologies (i.e., reference fish) were collected at an additional 22 stations and sent to Gulf Breeze. Upon examination by the Gulf Breeze specialists, only 8 (4.1%) of the 197 fish identified by the field crews as having at least one pathology of any type were found to lack the condition. This low overall error rate suggests that field crews were successful in identifying "diseased" fish (i.e., fish having one or more pathologies).

The field crews also were reasonably successful in identifying healthy fish; only 54 (9.4%) of the 572 reference fish were found by the Gulf Breeze specialists to have one or more pathologies. The presence of parasites in the branchial chambers of reference fish was the pathology most frequently missed in the field. A microscope normally is required to detect branchial parasites; therefore, field crews cannot be held accountable for these errors. The

field crews were negligent, however, in identifying fin erosion on 7 reference fish and gill erosion on 10. These results will be used to focus training efforts in future years.

Although field crews generally were successful in identifying both diseased (n = 197) and healthy fish (n = 572), an analysis of the QC data for all fish (n = 769) on the basis of individual pathologies revealed relatively high error rates for certain disorders. Two types of identification error were possible in the field: 1) identification of pathologies that were not confirmed by the Gulf Breeze experts (false positives), and 2) failure to identify pathologies that were identified at Gulf Breeze (false negatives). As indicated previously, the highest rate of false negatives occurred for branchial chamber parasites, for which the field technicians were not responsible. Pathologies that should have been identified but were missed frequently by the field crews included body lumps/bumps, skin discoloration, and erosion of fins and gills. Interviews with crew members suggested that crews became less diligent about looking for additional abnormalities after detecting a single pathology. False negatives, therefore, reflected a lack of effort on the part of field technicians rather than an inability to recognize common pathologies. In the future, crews will be trained to identify all pathologies on each fish. The field technicians also exhibited relatively high rates of false positive identifications for several pathologies, including body ulcerations, fin and gill erosion, skeletal malformation, and microphthalmia. Most of these misidentifications occurred on fish that, according to the Gulf Breeze experts, exhibited a different pathology than the one identified in the field.

Even though the identification of specific pathologies was often incorrect, the field crews achieved a low overall error rate for identification of "diseased" fish (i.e., fish exhibiting one or more pathologies of any kind). The high rate of false positives probably reflects insufficient training of field technicians in distinguishing pathologies. The 1990 QC results will be used to improve training in future years. Although the 1990 data were useful for making reliable estimates of the number of diseased fish in the Virginian Province, it was not possible to quantify the prevalence of specific pathologies with certainty.

The rocking chair dredge was deployed successfully at 104 of the 126 scheduled stations (Table 3-1) in an attempt to collect large (greater than 2.5 cm) bivalves. Although this indicates the logistical feasibility of deploying the equipment, the overall success rate for collecting bivalves was relatively low. Bivalves were collected and saved for chemical analysis at only 36 (29%) of the expected stations. Often, no bivalves were collected in areas where they were expected to occur. Although the dredge appeared to be operating properly, it often became clogged with mud and was ineffective. Extensive modifications of this gear are needed if collecting bivalves continues to be of interest to the program in future years. Sufficient biomass for chemical analysis of bivalve tissue was obtained from only 13 (10%) of the expected number of stations. Although these samples were analyzed, the data were insufficient to assess the utility of bivalve tissue contamination as an indicator, and it could not be used in the assessment of estuarine condition.

SECTION 4
INDICATOR DEVELOPMENT AND TESTING

EMAP monitoring focuses on four types of ecological indicators (Hunsaker and Carpenter 1990): response, exposure, habitat, and stressor (Fig. 2-4). Response indicators are ecological characteristics that integrate the responses of living resources to individual pollutants or combinations of pollutants and other stresses and provide quantitative evidence of the condition of ecological resources (Messer 1990). Exposure indicators quantify pollutant exposure and habitat degradation and are used mainly to identify probable causes of poor environmental quality that may explain observed responses of ecological indicators. Exposure indicators and response indicators are used to describe environmental status. Habitat indicators provide basic information about the natural environmental setting and are used to normalize exposure and response indicators to natural environmental gradients. Stressor indicators are used to quantify pollution inputs and stresses and identify the likely sources of pollution exposure.

This approach differs from many historical monitoring and assessment programs in that it emphasizes biological indicators for conducting ecological assessments (Hunsaker and Carpenter 1990). The biological indicators are selected based on their known responsiveness to anthropogenic pollutants, habitat modifications caused by human activities, and other influences believed to be causing degradation. This ecologically-based approach is emphasized in EMAP because it can be applied to situations where multiple stressors are acting separately or in combination and where natural processes affecting pollution exposure can not be modeled easily (e.g., the bioavailability of contaminants in sediments). This is certainly the case in estuarine systems, which are subject to an array of anthropogenic inputs and exhibit great biotic diversity and complex physical, chemical, and biological interactions. Using an ecologically-based approach may also reveal the effects of emerging environmental problems, even when causal relationships are poorly understood.

An important component of the 1990 Virginian Province Demonstration Project was an attempt to develop response indicators that can be used to discriminate between polluted and unpolluted environments on a regional scale. Although establishing biocriteria has been identified as an Agency priority, biotic indicators of ecological condition have not yet been developed for estuaries. The major reason is the lack of sufficient historical information with which to calibrate and verify the reliability of candidate response indicators over the large geographic and habitat gradients (e.g., salinity) that characterize estuaries. Rather than developing generic response indicators to estimate the degree of degradation, estuarine ecologists generally have relied on site-specific knowledge about differences in the kinds and abundances of organisms between polluted and unpolluted habitats and expert opinion to define the extent and severity of degradation. Because different kinds of organisms occur in different parts of a province, this approach is not appropriate on regional scales.

In contrast, methods for measuring the exposure variables selected by EMAP are well-developed and have been used in a number of monitoring programs throughout the province; however, some of the exposure variables, such as sediment toxicity, have not been applied on a regional scale. Others, such as dissolved oxygen, have been measured on regional scales but never using a probability-based approach to develop regional status estimates of the kind being made in EMAP.

Recognizing that measuring indicators on regional and national levels differs from the way such information has been used in many previous programs, EMAP developed a strategy to select and incorporate indicators into the program (Knapp et al. 1990; Holland 1990). In this strategy, potential indicators are incorporated into the measurement program if they meet a series of evaluation criteria that apply to all resource groups in EMAP (Hunsaker and Carpenter 1990). This section describes analysis of data collected during the 1990 Demonstration Project to determine how well some of the potential indicators for estuaries met the EMAP evaluation criteria.

4.1 RESPONSE INDICATORS

As previously described, response indicators are characteristics of the environment that provide quantitative evidence of the status of ecological resources and the biological integrity of the site at which they are measured (Messer 1990). Ecosystems with a high degree of biotic integrity (i.e., healthy ecosystems) are composed of balanced populations of indigenous organisms with species composition, diversity, and functional organization comparable to natural habitats (Karr and Dudley 1981; Karr et al. 1986). Response indicators include measurements of the kinds and abundances of biota present, the health of individual organisms, and the sustainability of critical ecological processes. They are the empirical data collected in EMAP that are integrated to measure the status and trends in the biological integrity of ecological resources.

Two categories of response indicators were measured during the 1990 Demonstration Project: (1) benthic response indicators, and (2) fish response indicators. Benthic organisms are invertebrates that live in the sediments of aquatic habitats. In estuaries, they are a major link between primary producers and higher trophic levels, including fish, shellfish, birds, and other wildlife (Carriker 1967; Rhoads 1974). They are a particularly important source of food for juvenile fish and crabs (Chao and Musick 1977; Bell and Coull 1978; Holland et al., 1989). Estuarine benthos also play important roles in ecological processes that affect water quality and productivity. For example, the feeding and burrowing activities of macrobenthos affect sediment depositional processes and chemical transformations (Carriker 1967; Rhoads 1974; Kemp and Boynton 1981). Benthic feeding activities also remove large amounts of particulate material from some shallow estuaries, which may improve water clarity (Cloern 1982; Officer et al. 1982; Holland et al. 1989).

Benthic assemblages have many attributes that make them reliable and sensitive indicators of the ecological condition of estuarine environments (Carriker 1967). For example, most macrobenthic species have limited mobility and cannot avoid exposure to anthropogenic or natural stress. Benthos live in bottom sediments, where exposure to contaminants is highest. Benthic assemblages are composed of a diverse array of species that respond to pollution stress in a variety of ways. Some species are especially sensitive to exposure to pollution. These species experience adverse effects due to this exposure, such as mortality, reduced growth, and decreased reproduction. Others are tolerant of pollution, and respond by increasing in abundance following exposure to pollutants. Because they generally depend upon and interact with biological processes in the water column, the responses of benthic organisms are representative of overall ecosystem responses (Rhoads 1974; Boesch and Rosenberg 1981; Holland et al. 1988).

Fish have several advantages as potential indicators of estuarine condition. Because fish have long life-spans and dominate the upper end of the food web, their responses integrate many short-term and small-scale environmental perturbations. They are known to respond to most of the major environmental problems of concern in estuaries (NOAA 1988), including eutrophication, habitat modification, and the presence of pathogenic or toxic contaminants. For example, eutrophication can affect fish adversely by diminishing dissolved oxygen concentrations below critical levels for growth or survival. Habitat modification, such as the loss of submerged aquatic vegetation, has been linked to decreased fish productivity through loss of important nursery areas and shelters from predation. Toxic and pathogenic contaminants can decrease fish growth, reproduction, or survival and can make fish unsafe for human consumption.

Fish also are valuable as indicators because of their importance in determining public perception of estuarine quality. Reports of fishery closures due to chemical and viral contamination alarm the public. In addition, the public has an economic interest in estuarine and coastal fisheries. More than seven billion dollars are spent annually in this country on saltwater recreational fishing, the vast majority of which occurs in estuaries or within three miles of the coast. Combining recreation and commercial fishing, estuaries and coastal waters account for 70% of U.S. fisheries landings (NOAA 1987).

4.1.1 General Methodology for Developing Response Indicators

The approach for developing response indicators consisted of five steps (Fig. 4-1):

1) developing a list of candidate measures for discriminating among sites of differing environmental quality,

2) identifying a set of sample sites that could be categorized with confidence as "nondegraded regional reference sites" or "degraded sites with known pollution exposure,"

3) using discriminant analysis to identify combinations of candidate measures that distinguished reliably between the degraded and nondegraded sites identified in step 2,

4) validating the index, and

5) scaling the index from 0 to 10.

The activities that occurred in each step during planning for the 1990 Demonstration Project are described below.

Step 1: Identify candidate measures

The first step consisted of two activities: (1) identifying candidate measures for differentiating between nondegraded reference sites and degraded sites, and (2) characterizing and removing variation in candidate measures associated with variation in natural habitat factors.

A list of candidate measures was developed by reviewing the relevant scientific literature and conducting a series of workshops attended by ecologists knowledgeable about northeastern U.S. estuaries and pollution assessment techniques. Measures that were applicable across multiple habitats and broad regions (e.g., measures of species richness) were sought. Measures were selected to represent the range of ecological attributes that characterize estuarine assemblages, including biodiversity, abundance, species composition, trophic interactions, and health of individual organisms. Measures that were applicable to a narrow range of habitats (e.g., salinities) or particular subregions were not included on the list.

A dominant feature of estuarine environments is the large spatial variability in physicochemical conditions that occurs over relatively small distances. For example,

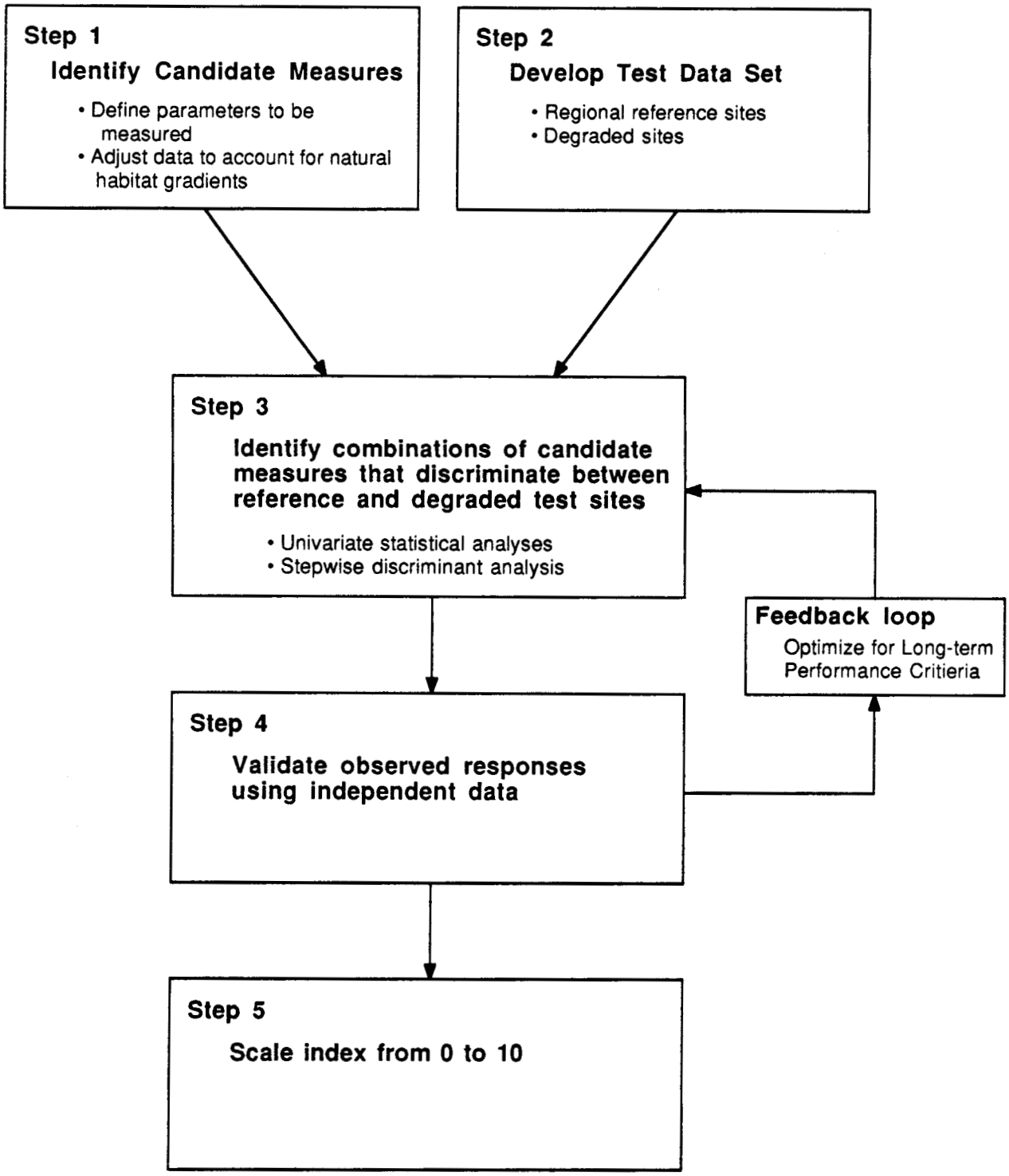

Figure 4-1. General methodology used to evaluate potential response indicators.

salinity variation that occurs from the headwaters of estuaries to their seaward boundary is a major factor controlling the biodiversity and abundance of estuarine biota (Carriker 1967; Boesch 1973; Lippson et al. 1979). Such large natural variation can obscure the responses of candidate measure to pollution exposure and must be identified and controlled for before the responses of candidate measures to pollution exposure can be characterized accurately. Procedures for determining the influence of natural environmental factors on candidate measures involved: 1) examining scatter plots of the distribution of candidate measures (Y axis) against habitat factors, including salinity, water depth, sediment silt-clay content, and latitude (X-axis); and 2) conducting linear or log-linear regressions to measure the magnitude, direction, and significance of relationships identified from scatter plots. Candidate measures were considered to be significantly influenced by habitat factors if the correlation coefficient (r^2) in regressions was greater than or equal to 0.25, and the slope was statistically different from 0. Regressions with correlation coefficients less than 0.25 were considered biologically insignificant. Data collected at all of the EMAP sites were used for these analyses. Candidate measures were corrected for the influence of habitat factors by reformulating them in a manner that was insensitive to habitat conditions (see Section 4.1.2)

Prior to performing regression analyses or other statistical examinations of the data, the distribution of each candidate measure was tested for normality and homogeneity of variance. Measures that were not normally distributed or had unequal variances were re-examined for assumptions of normality after making logarithmic, arcsin, and square root transformations. In all cases where adjustment was necessary, the Log_{10} (value +1) transformation was found to be an appropriate transformation.

Step 2: Develop a test data set

Step 2 identified EMAP sites that could be used to determine which candidate measures and what combinations of measures were most effective for discriminating among degraded and nondegraded conditions. Ideally, nondegraded reference sites should have been established by identifying pristine estuaries that are free of anthropogenic influence and selecting sampling sites within them. Such estuaries, however, could not be identified in the highly populated and heavily industrialized northeast United States. Instead, the reference sites were areas believed to be "least impacted" on the basis of known exposures, and degraded sites were those known to have substantial exposure to anthropogenic stresses. A preliminary list of such sites was established during the planning process (Fig. 2-3) based on historical water and sediment quality data, and the knowledge of local scientific experts.

Despite considerable effort to select appropriate test sites, the expected environmental conditions were not verified by exposure measures at almost half of the preselected sites (e.g., some supposedly nondegraded reference sites had high concentrations of contaminants in sediment, and the sediment was toxic to biota in laboratory tests). The data collected at

randomly selected base sites, however, provided a large database that was used for identifying additional sites that could be included in the indicator testing program. These additional sites ensured that the indicator testing sites represented the full range of environmental conditions occurring in the Virginian Province. Inclusion of these additional sites in the indicator testing data set also ensured that the number of samples was sufficient to calibrate indicators with an acceptable level of precision.

Data for sediment contaminant concentrations, sediment toxicity, and dissolved oxygen measurements (particularly the continuous measurements) were used to document environmental conditions at indicator test sites. These data were used to select indicator test sites representing the

- range of habitats that exists in the Virginian Province, including all major salinity zones, sediment types, and biogeographical divisions;
- major categories of pollution stress, including contaminated sediments, stressful low dissolved oxygen concentrations, and the cumulative impact of stressful low dissolved oxygen concentrations and contaminated sediments; and
- relatively unpolluted reference sites that were exposed to little anthropogenic perturbation.

To establish a matrix of test sites, sampling locations were classified as nondegraded regional reference sites if:

1) summertime bottom DO measurements were never less than 1 ppm, 90% of the dissolved oxygen observations were greater than 3 ppm, and 75% of the dissolved oxygen observations were greater than 4 ppm;

2) no contaminant was apparent in the sediment at a concentration higher than that identified by Long and Morgan (1990) as the median effects concentration (ER-M value) for biological responses; and

3) survival of the amphipod *Ampelisca abdita* in the 10-day acute sediment bioassay was greater than 75% and not significantly different from controls.

Sites were classified as degraded if:

1) dissolved oxygen concentrations below 0.3 ppm were recorded at the site during the sampling period, or if 10% of the continuous bottom dissolved oxygen observations was less than 1 ppm, or 20% of bottom dissolved oxygen

observations was less than 2 ppm, or bottom dissolved oxygen observations of less than 2 ppm occurred at the site for 24 consecutive hours; or

2) the concentration for at least one sediment contaminant exceeded the ER-M value defined by Long and Morgan (1990), and the survival for a 10-day acute sediment bioassay using the amphipod *Ampelisca abdita* was less than 75% of, and significantly different from, control survival.

These criteria do not represent any EPA standards and are intended only for use in this report.

Application of the above criteria identified a subset of sample sites that probably were classified accurately as relatively undisturbed regional reference areas and degraded sites with known pollution exposure. The number of sites identified as indicator test sites was, however, considerably fewer than the total number of sites sampled. This is because of the conservative nature of the criteria used to assign sites to categories. This conservative procedure was necessary because of the poor scientific understanding of biotic responses to intermediate levels of pollution stress.

Degradation resulting from habitat modification and other causes was not included in defining indicator test sites. The data required to document the degree of habitat modification at a site quantitatively are not available currently, and the resources were not available to collect these data. The responses of candidate response indicators to habitat modification and other causes of degradation were assumed to be similar to the responses of the pollution stresses tested.

Regional reference sites were selected to calibrate the responses of indicators identified in step 1 into an index that defines biological condition. Defining biological condition relative to regional reference sites believed to contain nondegraded resources has a two important implications with respect to interpretating the resulting index. First, degraded and nondegraded were defined relative to present-day conditions. The conditions at present day reference sites may not represent nondegraded conditions that existed decades or centuries ago. Second, degradation determined by comparison to biological condition at reference sites does not lend itself to distinguishing degradation attributable to human activity from degradation due to natural processes. For example, the presence of low dissolved oxygen may cause a biological response regardless of whether the low DO is due solely to the presence of deep, highly stratified waters or results, in part, from anthropogenic enrichment of nutrients and organic carbon. Determining the relative contribution of natural and anthropogenic factors would require intensive, site-specific or system-specific studies of metabolic processes that are beyond the present scope of EMAP.

Step 3: Identify combinations of candidate measures that discriminate between degraded and nondegraded areas

In Step 3, discriminant analysis was used to identify combinations of candidate measures from the list generated in step 1 that reliably distinguished between the degraded and nondegraded sites identified in step 2. Discriminant analysis provides an unbiased tool for selecting index components from among the list of candidate measures. The approach also provides a means for weighting component measures that is free from investigation bias. Finally, an unbiased threshold value for distinguishing between sites having degraded and nondegraded condition could be defined using discriminant analysis.

Prior to conducting discriminant analysis, however, a t-test was used to define the direction and magnitude of differences for each candidate measure. Only candidate measures for which there was a significant ($p<0.05$) difference in parameter values between reference areas and degraded sites were included in the discriminant analysis. This procedure prevented including measures with high signal to noise ratios. It also reduced the list of candidate measures to a manageable number from which it was highly probable that a subset(s) could be identified to discriminate reliably between reference and degraded areas.

In the stepwise discriminant analysis, a p-value of 0.15 was the criterion for including a measure into the model. In addition, the direction of the coefficient for each measure included was required to be consistent with the direction of the difference observed in the t-test. If the direction of a measure was inconsistent between the t-test and discriminant analysis, the stepwise procedure was repeated after removing the measure in question. This procedure reduced the possibility of identifying a combination of measures that optimized only for the calibration data set. Canonical discriminant analysis was used to determine the direction and magnitude of coefficients.

Many of the candidate measures contain redundant information (i.e., are highly correlated); thus, although the combination of measures selected by discriminant analysis maximizes separation between degraded and reference sites, it may fail to incorporate easily interpreted and highly valued measures (e.g., species richness) that also would be useful for discriminating among sites of differing quality. For this reason, the discriminant procedures were applied twice. The first analysis was conducted allowing the analysis procedure to include measures only on the basis of the degree to which they differentiated between degraded environments and nondegraded regional reference areas. In the second discriminant analysis, selected measures were forced into the model. These measures were not included in the first model, but were reported in the scientific literature to discriminate between degraded and nondegraded environments (e.g., species richness). Further, the F-values and partial r^2 from the first analysis indicated these forced measures were highly correlated with measures that were included. This iterative process identified several combinations of candidate measures that distinguished between degraded and nondegraded sites. For each of these models, canonical

discriminant analysis was used to calculate the frequency with which reference sites were incorrectly classified as degraded (i.e., false positive), and the frequency with which degraded sites were classified as reference areas (i.e., false negatives).

As a means for maximizing discrimination among two-predefined groups, stepwise discriminant analysis is sensitive to the incorrect assignment of sites. To evaluate the degree to which individual stations influenced the discriminant analysis results, canonical discriminant analysis was conducted multiple times. Each time, a different station was eliminated from the calibration data set. Stations that had a large influence on discriminant coefficients were identified and examined to determine if they were incorrectly assigned to test groups. Stations that were judged to be misclassified were removed from the test data set and step 3 was repeated.

Step 4: Validating the index

Step 4 involves validating the indices developed in the previous steps. Validation requires testing the index with an independent data set to ensure that the multivariate solution is not specific to the indicator testing sites sampled in 1990. Validation will be accomplished with data collected in future years of the program; however, it was possible to validate the indices preliminarily in two ways with the 1990 data: 1) the cross-validation procedure of Lachenbruch (1975), in which each station is removed from the calibration data set and used as a test-case for validation, and 2) evaluating the performance of the indices using data collected at 16 of the test sites that were revisited in interval 3 (September, 1990). Period 3 data are not entirely independent because they were collected at the same sites as the period 2 data; nonetheless, these data do permit examination of consistency of the response, which is a necessary attribute of an index.

Step 5: Scaling the index

Scores from canonical discriminant analysis are unscaled and are difficult for general audiences to understand; therefore, discriminant scores were normalized to a scale of 0 to 10 using the following algorithm:

4.1.2 Benthic Macroinvertebrates

4.1.2.1 Development of a Benthic Index of Estuarine Integrity

Characteristics of benthic assemblages have been used to measure and describe ecological status and trends of marine and estuarine environments for the last several decades (Sanders 1956; Sanders 1960; Rosenberg 1976; Boesch 1973; Pearson and Rosenberg 1978; Rhoads et al. 1978; Boesch and Rosenberg 1981; Holland et al. 1988). This literature has identified a diverse array of measures of benthic assemblages that represent ecological status and trends, including 1) measures of biodiversity/species richness; 2) changes in species composition; 3) changes in the relative abundance or productivity of functional groups; 4) changes in the relative abundance and/or productivity of "key" species; 5) changes in biomass; and 6) relative size of biota. Estuarine benthic ecologists have used site-specific knowledge about differences in the kinds, abundances, and physiological condition of individual benthic organisms between degraded sites and undegraded reference sites to estimate the extent and degree of degradation. Although many different attributes of benthic assemblages have been identified and used previously to measure ecological status and trends, a generally accepted (i.e., calibrated and validated) benthic indicator that integrates the information available from all the potential measures into a single measure (i.e., a benthic index value) that is applicable over broad geographical areas and sensitive to a broad range of stresses has not been developed previously.

Differences in the specific attributes of the benthic assemblage measured, collection methods used, timing of sample collections, and specific objectives of programs limit the usefulness of the existing data for developing a benthic index. For example, the level of taxonomic identification and size of organisms sampled frequently varies from study to study, making the data inappropriate for developing a benthic index. Data used to develop a benthic index must be collected in a standardized way over broad geographical scales. Most of the existing benthic data do not meet these criteria and cannot be used to develop a broadly accepted benthic index.

Although the historical data for benthos cannot be used to develop a benthic index, these studies provide valuable information that was used to design the EMAP-E benthic indicator testing program. Historical data was used to identify candidate measures for inclusion in the measurement program and develop cost-effective sampling and processing methods. Several indicator workshops were held to review and refine the proposed benthic indicator testing program. A goal of the Demonstration Project was to collect the data and conduct the analyses necessary to develop a benthic index that would produce information that is understood easily by technical and nontechnical audiences.

Step 1: Identify candidate benthic measures

Benthic abundance, biomass, and species composition data were used to define 22 descriptors of the major ecological attributes of the benthic assemblages occurring at each sample site (Table 4-1). Many of these descriptors were formulated in several ways. The measures of taxonomic composition and functional groups generally were calculated on the basis of both biomass and counts. In addition to absolute values, measures also were expressed as proportions of total abundance or biomass, where appropriate. Some of the measures, such as biodiversity, were calculated by averaging and by compositing the three replicate samples. When the 22 descriptors were reformulated in these ways, 58 candidate measures were defined.

The 1990 Virginian Province Demonstration Project sampled only benthic organisms that do not pass through a 0.5-mm mesh sieve. This component of the benthos constitutes greater than 90% of the biomass and is relatively stable over long periods of time (Mare 1942; Rhoads 1974). Only data for infauna (organisms that live in or on the sediment surface) were used when computing values for candidate measures. Data for epifauna (organisms that live on hard surfaces such as shells) were excluded from computations because 1) the sampling methods did not sample epifauna efficiently; 2) the exposure of epifauna to pollution insults, particularly chemical contaminants in sediments, is different than the infauna; and 3) shell and other hard bottom habitats were found at relatively few sample sites (less than 1%). Excluding epifauna reduced the possibility that the presence of small pieces of shell or other structure that occurred at some sites would introduce unnecessary variability or bias analysis results. Over 90% of the individuals collected were infauna.

Almost all of the candidate benthic measures were significantly ($p<0.05$) correlated with at least one of the habitat factors measured. Over 75% of the candidate measures were related significantly to salinity distributions (Table 4-2). Relationships between candidate measures and other habitat factors (i.e., latitude, silt-clay content of sediments, water depth) occurred less frequently and did not account for as much of the total variation as relationships with salinity (Table 4-2).

Correlations between candidate measures and habitat factors, however, explained little of the total variation in candidate measures. Only two of the correlations accounted for more than 25% of the total variation. Both of these were measures of species richness: 1) mean number of species per site, and 2) total number of species per event. Salinity accounted for up to 35% of the total variation for these measures. Figure 4-2 shows the distribution of the number of species per site in response to salinity.

Table 4-1. Test for significant differences between degraded and nondegraded reference sites for each of the descriptors of the benthic assemblage. P-values are shown for the formulation of each descriptor that best distinguished sites of different quality.

	t-test (p-value)	Direction (+ = greater value at reference sites)
Measures of Biodiversity		
Shannon-Weiner index	0.06	+
Pielou's evenness index	0.18	+
Proportion of expected number of species	< 0.001	+
Measures of Community Condition		
Total benthic biomass	0.03	+
Total benthic abundance	0.005	+
Measures of Individual Health		
Biomass/abundance ratio	0.52	+
Weight per individual polychaete	0.02	+
Weight per individual mollusc	0.97	0
Weight per individual bivalve	0.98	0
Measures of Functional Groups		
Suspension feeding species abundance	< 0.001	+
Deposit feeding species abundance	0.02	+
Omnivore/predator species abundance	0.001	+
Opportunistic species abundance	0.16	-
Equilibrium species abundance	0.001	+
Measures of Taxonomic Composition		
Amphipod abundance	< 0.001	+
Bivalve abundance	< 0.001	+
Gastropod abundance	0.02	+
Molluscan abundance	< 0.001	+
Polychaete abundance	0.02	+
Capitellid polychaete abundance	0.006	+
Spionid polychaete abundance	0.9	+
Tubificid oligochaete abundance	0.49	+

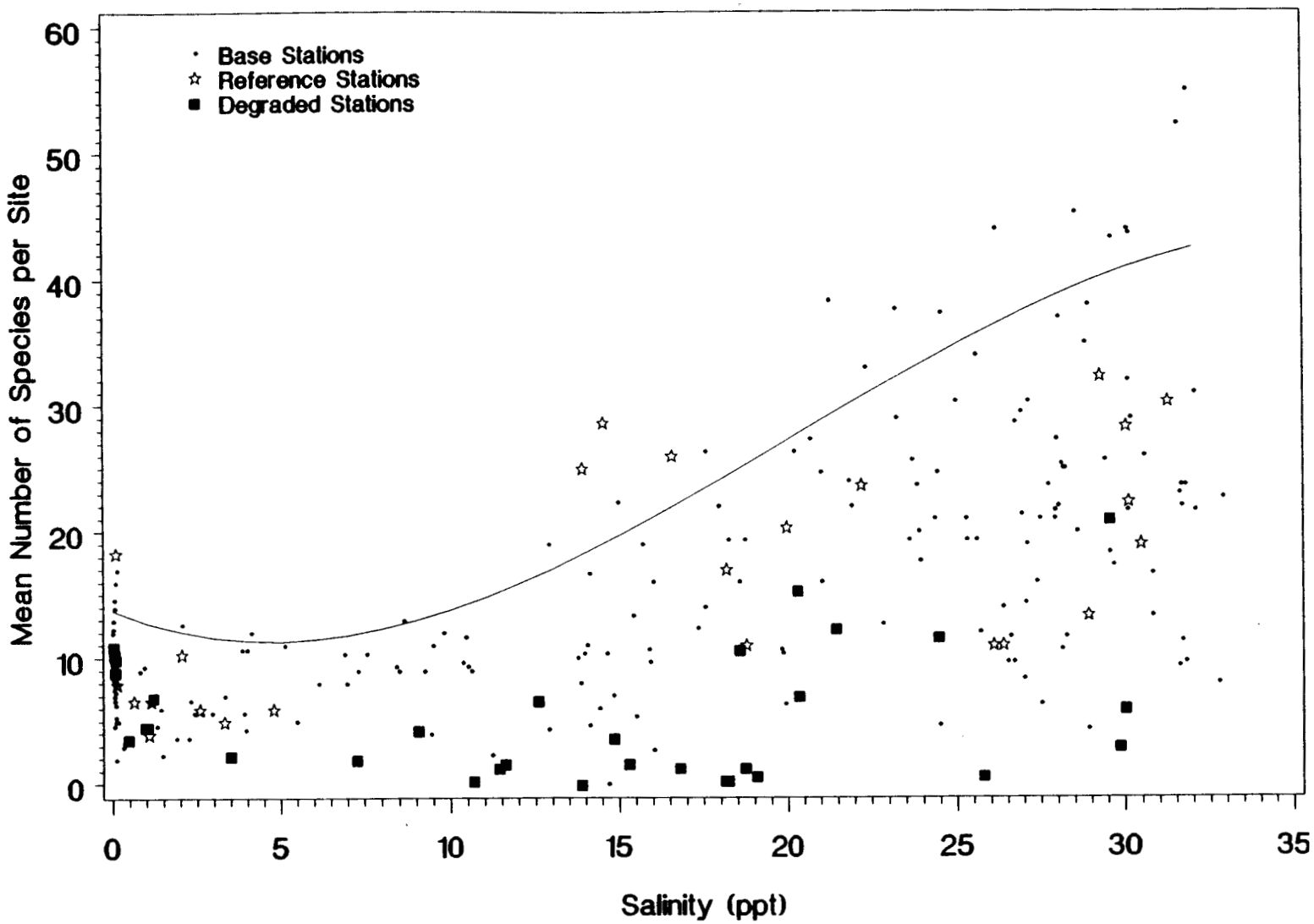

Figure 4-2. Plot showing the regression line used to estimate salinity-adjusted species richness measures for mean number of species per site. The distribution of reference and degraded sites relative to the estimated line is shown. The line is the expected number of species based on the polynomial regression.

Table 4-2. Summary of correlation between habitat indicators and the candidate benthic measures			
Habitat Indicator	**Number of Significant Correlations**	**Number of Correlations with $r^2 > 0.10$**	**Number of Correlations with $r^2 > 0.25$**
Salinity (ppt)	43	12	2
Latitude	37	2	0
Silt-clay Content	23	5	0
Water Depth	19	3	0
All	55	18	3

The two measures of benthic species richness were corrected for the effects of salinity by calculating adjusted measures of species richness that were insensitive to salinity variation. The adjusted measures consisted of calculating the percent deviation in species richness from a baseline condition that represented the response of benthos to the estuarine salinity gradient. The adjusted measurement was renamed percent expected number of species. The baseline for estimating the percent expected number of species was obtained by fitting a polynomial through the 90th percentile of a 3 ppt running average of species richness measures using all the data collected in 1990. The r^2 for this relationship was 0.98. The 90th percentile was chosen because it did not differ substantially from the actual species richness values observed for nondegraded reference sites and represented the number of species that would be expected to occur if the only factor influencing species richness measures was the estuarine salinity gradient. Figure 4-3 is the scatter plot showing the insensitivity of percent expected mean number of species per site to salinity. This measure was also insensitive to sediment type, depth, and latitude. Baselines based on other percentiles (75th, 50th) had similar distributions and would have produced similar results.

The scientific literature suggests that the relative abundance of benthic organisms frequently is related to sediment characteristics (e.g., percent fine sediment particles); however, only five of the candidate measures had relationships with the silt-clay content of sediments that accounted for more than 10% of the total variance (Table 4-2). All of these were negative relationships with species richness measures, suggesting that mud habitats are inhabited by fewer species than sand habitats. These variations were not

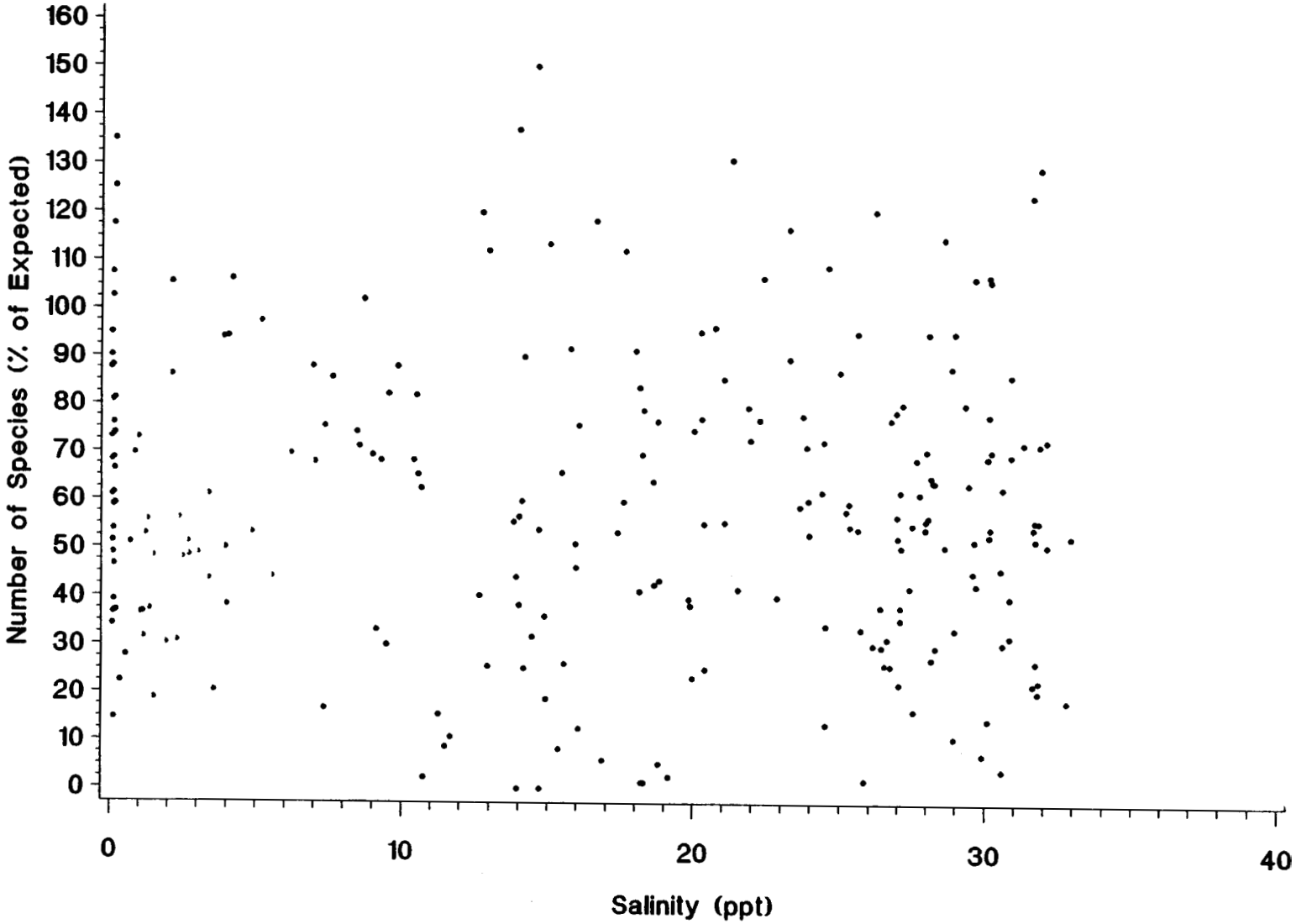

Figure 4-3. Relationship of number of species to salinity after adjusting for habitat variables.

corrected because these relationships did not account for a significant proportion of the variability in the data. In addition, mud sites had the greatest exposure to low dissolved oxygen and contaminant stress. The mud habitats represented by lowest species richness and diversity values were the same stations exposed to most severe low dissolved oxygen and contaminant stress. Adjusting for this variation would have decreased the probability of accurately representing the responses of benthic assemblages to pollution.

Step 2: Develop a test data set

Thirty-three sites were selected from base and ITE sampling sites for use in the indicator testing data set; 14 of these were nondegraded reference sites (Table 4-3). The 19 remaining sites were classified as degraded based on the criteria established for low dissolved oxygen stress, exposure to contaminants, or both. Twenty-seven percent of the test sites were north of the Hudson River. Both the degraded and nondegraded test sites encompassed a wide range of habitats (Table 4-4), including all major salinity zones and sediment types that occur in the Virginian Province; however, no test sites were located in oligohaline sand. Mud substrate was more common at degraded test sites than at reference sites, which probably reflects the affinity of contaminants for fine-grained particles.

Step 3: Identify combinations of candidate benthic measures that discriminate between degraded and nondegraded areas

Twenty-eight of the 58 candidate benthic measures differed significantly between degraded and reference sites in the t-tests (Table 4-1). These differences were in concordance with expectations established when the measures were proposed (e.g., total abundance, abundance of sensitive groups, and species richness measured were higher at regional reference sites than degraded sites). These 28 measures were the only candidates used in the discriminant analyses.

Discriminant analysis results are summarized in Table 4-5. Four variables were included in the model for the first stepwise discriminant analysis (Index 1): 1) number of amphipods per event, 2) the percent of total benthic abundance in molluscan taxa, 3) the mean weight per individual polychaete, and 4) the number of capitellids per event. Using this combination of measures, 7% of the nondegraded reference sites were classified as degraded (i.e., false positives), and 11% of degraded sites were classified as reference sites (i.e., false negatives) (Table 4-5). The canonical r^2, which approximates the total variance explained by this analysis, was 60%.

Table 4-3. List of sites used in indicator testing dataset sampled during the 1990 Virginian Province Demonstration Project

Habitat Class	Reference Sites	Degraded Sites		
		Low Dissolved Oxygen Stress	Contaminant Stress	Low Dissolved Oxygen and Contaminant Stress
Extreme low salinity habitats (< 5 ppt)	Hudson River (101)* 41°23'00" 73°57'23" Rappahannock River (192) 37°57'54" 76°52'02" Elk River (254) 39°28'47" 75°56'30" Bohemia River (89) 39°22'42" 76°00'00" Potomac River (186) 38°30'00" 77°16'29" Potomac River (188) 38°44'12" 77°02'00"	None	Housatonic River (169) 41°17'12" 73°04'19" Delaware River (223) 39°45'00" 75°29'00"	Anacostia River (88) 38°52'11" 76°59'51" Back River (90) 39°16'12" 76°26'36"
Brackish habitats (5-18 ppt)	Tangier Sound (41) 38°01'41" 75°54'06" Barnegat Bay (256) 39°56'36" 74°06'07" Chesapeake Bay (57) 37°26'04" 76°14'07"	Chesapeake Bay (62) 38°59'12" 76°21'29" Chesapeake Bay (65) 38°33'27" 76°24'05" Potomac River (180) 38°04'13" 76°27'53" Potomac River (182) 38°13'06" 76°47'08"	Colgate Cove (82) 39°15'12" 76°33'06" Passaic River (103) 40°45'00" 74°09'54"	Bear Creek (81) 39°14'36" 76°29'29" Patapsco River (134) 39°14'47 76°33'25"
Estuarine habitats (> 18 ppt)	Delaware Bay (33) 39°12'36" 75°12'43" Indian River Bay (150) 38°35'36" 75°06'42" Napeague Bay (159) 41°03'42" 72°00'06" Sandy Hook Bay (122) 40°27'36" 74°04'30" Narragansett Bay (70) 41°38'29" 71°18'01"	Chesapeake Bay (56) 38°08'40" 76°14'05" Chesapeake Bay (80) 38°53'23" 76°24'04" Mystic River (106) 41°21'53" 71°57'52"	Arthur Kill (94) 40°37'18" 74°12'12" Raritan Bay (260) 40°42'17" 74°06'59"	Elizabeth River (86) 36°49'55" 76°17'38" Blackrock Harbor (98) 41°09'35" 73°12'37" New Bedford Harbor (99) 41°38'33" 70°54'42"

* Number in parentheses represents EMAP station number.

Table 4-4. Number of indicator testing sites sampled in each of several habitat types		
Habitat Type	**Number of Sites**	
	Degraded	**Nondegraded**
Sand		
Oligohaline	0	0
Mesohaline	2	1
Polyhaline	1	3
Mud		
Oligohaline	4	6
Mesohaline	5	0
Polyhaline	7	4

Changes in the F-statistic and partial r^2 at steps 1 and 2 of the first discriminant analysis indicated that measures of amphipod abundance were correlated highly with salinity- adjusted measures of species richness (percent expected number of species); therefore, in the second discriminant analysis, the percent expected number of species was forced into the model. Five measures were included in the second discriminant model (Index 2): 1) percent expected number of species, 2) the percent of total benthic abundance as amphipods, 3) the percent of total benthic abundance in molluscan taxa, 4) the mean weight per individual polychaete, and 5) the number of capitellids per event. Classification efficiency with the salinity-adjusted species richness measure included was the same as in the previous analysis, and the canonical correlation coefficient was slightly higher (Table 4-5).

Sensitivity analysis indicated that two sites were having a disproportionate effect in determining calibration of both discriminant models. These two sites (94 and 106), both of which were classified as degraded sites, were examined more closely to determine whether their disproportionate effect potentially resulted from misclassification of the sites. Site 94 is located in the Arthur Kill estuary in northeastern New Jersey, and the data used to classify the station as degraded were compelling: 1) 99% of the amphipods in the bioassay died; 2) several metals and organic chemical concentrations exceeded Long and Morgan's (1990) ER-M values; and 3) historic water quality data indicated that the site frequently experienced hypoxic conditions in summer months. Based on this information, it was concluded station 94 was classified correctly and should remain in the calibration data set.

Table 4-5. Results of discriminant analyses conducted to combine candidate benthic measures into an index

Analysis	Selected Measures	Calibration Efficiency		
		Percent of Degraded Sites Correctly Classified	Percent of Reference Sites Correctly Classified	Canonical R^2
Index 1 • No variables forced • All stations included	•Number of amphipods •Percent of individuals as molluscs •Mean weight per individual polychaete •Number of capitellids	89%	93%	.60
Index 2 • Species richness forced • All stations included	•Percent expected number of species •Percent of individuals as amphipods •Percent of total abundance as bivalves •Mean weight per individual polychaete •Number of capitellids	89%	93%	.62
Index 3 • No variables forced • Station 106 removed	•Number of amphipods •Percent of total abundance as bivalves •Mean weight per individual polychaete •Number of capitellids	89%	86%	.67
Index 4 • Species richness forced • Station 106 removed	•Percent expected number of species •Number of amphipods •Percent of total abundance as bivalves •Mean weight per individual polychaete •Number of capitellids	89%	86%	.67

Site 106 is located in the Mystic estuary, Connecticut, and reexamination of the exposure data for this site suggested it was probably misclassified. The sediment chemistry and sediment toxicity data did not indicate a contaminant problem at this site. The original basis for classifying this site as degraded was the occurrence of dissolved oxygen concentrations of 0 at the site. These extreme low dissolved oxygen readings were less than one hour in duration and occurred on only one day. Dissolved oxygen concentrations of less than 2 ppm were observed infrequently at site 106 during the rest of the summer (i.e., <6% of the time). Because low dissolved oxygen concentrations were rare at this site, discriminant analyses were repeated after removing station 106 from the calibration data set.

When station 106 was removed, the measures selected by the stepwise discriminant analysis remained basically the same as those selected previously (Table 4-5). For the analysis in which no variables were forced (Index 3), percent of total abundance in molluscan taxa was replaced with a similar measure, percent of total abundance as bivalves, and all other measures remained the same. For the analysis in which species richness was forced (Index 4), percent of total abundance as amphipods was replaced with total number of amphipods per event. Removing station 106 from the calibration data set also improved the canonical r^2 in both the forced and unforced discriminant analyses. The percent of sites classified correctly, however, declined slightly (Table 4-5).

Step 4: Validating the model

Four potential indices were developed in step 3. The cross-validation efficiency for all of them was about 80% (Table 4-6). The second means for validating indices was to determine the classification efficiency for 16 sites that were revisited in September. For the first two indices, 15 of the 16 sites were classified correctly, and for the third and fourth indices, 14 of the 16 sites were classified correctly. One site (#223), a degraded site located in the Delaware River, was misclassified by all four indices in the September data. This site was classified correctly by all indices using the August data. These findings suggest that small-scale spatial patchiness, or measurement error, rather than an inherent shortcoming in the index parameters, may have been responsible for the error associated with this site. All of the indices had correlation coefficients that exceeded 0.8 between August and September (Table 4-6).

Table 4-6. Results of validation of four candidate indices. Description of measures comprising each index is shown in Table 4-5.

Index	Cross-validation Efficiency		September Samples		
	Percent of Degraded Sites Correctly Classified	Percent of Reference Sites Correctly Classified	Overall Percent Correctly Classified	Index Correlation with August Samples	Percent of Sites Correctly Classified
1	84%	71%	79%	0.80	94%
2	79%	86%	82%	0.81	94%
3	83%	86%	84%	0.84	88%
4	83%	86%	84%	0.84	88%

Step 5: Scaling the index

The final step in developing the benthic index was to select one of the indices based on the calibration and validation information, calculate discriminant scores for all sample sites, and normalize the calculated scores to a scale of 0 to 10. Although all four candidate indices were calibrated and validated preliminarily at an acceptable level, the fourth alternative was used to assess the status of benthic resources of the Virginian Province. This index consisted of forcing species richness and dropping the questionable station from the calibration step. The discriminant function for this index was:

Discriminant Score =

(0.011 * Proportion of expected number of species +
(0.817 * Number of amphipods) +
(0.671 * Percent of total abundance as bivalves) +
(0.465 * Number of capitellids) +
(0.577 * Average weight per individual polychaete).

When applied to all 152 sites sampled, the range of this index was from -2.12 to +4.75 with a critical value for discriminating between degraded and reference sites (calculated as the midpoint between mean discriminant scores for degraded and reference sites) of +0.22. When these discriminant scores were normalized to a range of 0 to 10, the critical value between degraded and nondegraded was 3.40.

4.1.2.2 Discussion of the Benthic Index

Although the benthic index developed above appears to work well for distinguishing sites of differing environmental quality, it may not be the only effective index, or the most effective one. First, covariance among many of the candidate measures was high, indicating that many alternative combinations could produce comparable results. Second, index development was based on only 33 indicator testing sites that, although representative of a wide range of physicochemical and stress conditions, did not represent all possible conditions. In future development of the benthic index, additional sites will be added to the calibration data set so that it includes the full range of environmental habitats and stressors present. Third, there may be other important measures of the benthic assemblage that were not included as candidates in the stepwise discriminant analysis. As more is learned in local studies about other measures that are effective for discriminating sites of differing environmental quality, they can be incorporated into the calibrations. The index development process is flexible; as other studies suggest increased confidence in selected measures, they can be incorporated into the index through the forced stepwise discriminant analysis.

Although further index development is warranted, particularly validation with an independent data set, it is not clear that the predictive capability of the index will change dramatically. The process described above identified four possible indices. When they were applied to all sites for which there was benthic data, the correlation between the four index values exceeded 90% for all pairwise combinations (Table 4-7). Similarly, over 90% of stations in this data set were classified the same (degraded vs nondegraded), regardless of the index used.

Table 4-7. Correlation coefficients relating the values of each of four indices (top of table) and percent of sites that were classified the same (degraded vs. nondegraded) among four indices (bottom of table) applied to 152 sampling sites. Components of each index are provided in Table 4-5.

	Index 1	Index 2	Index 3	Index 4
Index 1		.97	.96	.96
Index 2	94%		.96	.96
Index 3	95%	91%		.99
Index 4	95%	91%	100%	

As development of a benthic index of estuarine condition continues, several questions need to be addressed. One of these is, "how good is good enough?" In the analyses described

above, 90% accuracy and about 84% cross-validation were the best classification efficiency rates that could be achieved. These may be acceptable because, even with a 10% to 15% error rate, biological data may be interpreted more easily than exposure information such as sediment chemistry alone.

The degree of control that investigators have for achieving a higher degree of accuracy may be limited. Some error may be due to sampling or measurement variability, as suggested by the fact that 9% of the sites switched classification from August to September. This error probably can be reduced by taking a larger sample or by enforcing greater quality control in the laboratory; however, doing so adds considerable cost. The sample size with three replicates used in the Demonstration Project already exceeds 1200 cm^2, and the quality control protocol required over 90% accuracy in sample sorting and identification. Some of the error may be due to the way sites in the calibration dataset were classified using dissolved oxygen and contaminant exposure. This error could be reduced by collecting additional measures, such as those that describe habitat alteration, but again at considerable cost. Finally, achieveable levels of sensitivity may be limited by the nature of the attributes being examined; assemblage measures may not be sensitive to some types of exposure. This limitation could be addresed by adding new types of indicators, such as measures of the sublethal responses of individuals, but these types of biomarkers are just beginning to be developed for invertebrates.

Several other factors that have not been addressed fully here need to be considered in further development of the benthic index. The index developed here may not be the most cost-effective. It is based on three replicate samples and incorporates biomass measures. It might be possible to reduce sampling and laboratory processing effort with little loss in classification efficiency. Additionally, the index was developed for a single region of the country. Although it might be the optimal solution for classifying sites in that region, a more general solution that is applicable nationally might be possible, with very little loss in effficiency at the regional level. The index development approach above, once augmented with additional data for other regions of the country and validation data, will allow these alternatives to be evaluated.

4.1.2.3 Sediment Profile Imaging Techniques

Traditionally, the condition of benthic assemblages is evaluated using quantitative measures of species composition, abundance, and, in some cases, biomass. An alternative approach to assessing benthic condition is sediment profile imaging, which involves still photography of the sediment-water interface to quantify various physical, chemical, and biological characteristics of sediments. These characteristics include sediment grain size, penetration of oxygen into the sediments, and a qualitative description of benthic community composition (Rhoads and Germano 1982, 1986). Profile imaging offers the advantage of rapid availability of data, as automated image analysis can provide data in nearly real time. Eliminating laboratory

processing also can reduce cost per sample, although capital costs associated with the equipment are substantial.

Although sediment profile imaging has been applied widely in monitoring dredged material disposal sites and organic enrichment gradients, it is not yet clear that the technology is developed sufficiently for application to a regional monitoring program such as EMAP, which involves sampling a wide variety of habitats. There have been only a few direct comparisons between sediment profile imaging and traditional benthic assessment methods, and these studies were conducted in fine-grained, polyhaline estuaries (Scott et al. 1987; Long et al. 1990; Grizzle and Penniman 1991). Although these studies support the use of sediment profile imaging in high salinity (greater than 15 ppt) waters, the technique has not been tested adequately in moderate or low salinity habitats (Holland et al. 1988, 1989), or in habitats with coarse-grained sediments. Application of such techniques in EMAP requires a thorough understanding of indicator responses across all habitat types.

During the Demonstration Project, sediment profile imaging was attempted at 20 sites where benthos were collected using traditional techniques. The objectives of this sampling were to determine the logistic feasibility of sediment profile image sampling across the broad range of habitat types and to compare the Organism-Sediment Index (OSI) values produced from the sediment profile images (see Section 3 for calculation of this index) with the benthic index (BI) values obtained from the same habitats.

Logistical difficulties with sediment profile imaging were encountered in sampling coarse grained sediments, which resulted in being unable to gather data from 15% of the test sites (see Section 4). At the remaining 17 sites, the OSI and benthic index were statistically correlated ($p<0.05$), but with a low correlation coefficient ($r^2 = 0.36$). Threshold values distinguishing degraded from nondegraded benthic conditions are 3.4 for the benthic index (this report) and 6 for the OSI (Valente et al. 1992). Ten of the 17 sites were classified in the same condition by both indices (Table 4-8). Six of the seven sites that were classified differently were classified as degraded by the OSI and nondegraded by the benthic index. Of these, three exhibited OSI values at or just below the threshold. Over half (nine) of all OSI values were at the threshold value, or one unit away from it, despite a total range for the index of 22 units (-11 to 10). Small shifts in the OSI, therefore, can dramatically change the classification analysis in this test.

The test was limited by the selection of sites at which profile imaging was conducted. Although a primary purpose of the study was to examine the technology in low salinity and coarse-grained environments, only three samples were collected in sites with salinity less than 15 ppt, and only three samples were collected from sites where silt-clay content was less than 20%. For two of the three low salinity sites, the OSI and benthic index classified the sites similarly. Of the three coarse-grained sites, the two methods classified one site differently.

Table 4-8. Comparison of benthic characterizations produced by sediment profile image analysis and conventional macroinvertebrate sampling. The threshold value for the benthic index (BI) is 3.4; the threshold value for the Organism-Sediment Index (OSI) is 6.

Station	Salinity (ppt)	Sediment Type (% silt/clay)	Sediment Profile Image (OSI)	Conventional Sampling (BI)
98	25	57	-8	0
215	0	91	5	2.4
95	27	14	6	2.5
177	7	90	-1	4.1
199	4	89	6	4.9
71	32	81	7	5.8
76	30	5	6	6
24	28	52	4	6.1
106	25	93	-3	6.1
25	28	87	6	6.6
26	27	74	9	6.9
21	28	91	4	7.8
74	32	15	10	7.8
73	32	63	5	8.1
70	31	56	7	8.5
96	27	93	5	9.4
7	28	24	9	10

The test was limited further because sediment profile imaging was not conducted at sites of known quality, such as those in the indicator testing data set; thus, it is not possible to determine which of the two methods provided a more accurate assessment of local conditions. The uncertainty associated with sediment profile imaging in selected habitats, combined with logistical difficulties associated with collecting samples in those habitats, indicate that replacing the conventional benthic invertebrate sampling with sediment profile imaging presently is not an appropriate option for EMAP. Further testing of the technology appears warranted since the available data are limited and insufficient to conclude that the technology is not feasible; moreover, recent advancements in reducing the size of profile imaging equipment may reduce the cost and eliminate some of the logistical obstacles associated with sampling coarse-grained sediments.

4.1.3 Fish

4.1.3.1 Development of a Fish Index of Estuarine Quality

Indices of environmental health based upon the fish assemblage occurring at a site have been accepted widely for use in freshwater environments. The Index of Biotic Integrity (IBI; Karr 1981) has become a standard measure for defining environmental quality in several states (Plafkin et al. 1989). To date, a validated index of estuarine environmental quality based on information about fish assemblages that is applicable over broad geographic areas has not been developed. During the Demonstration Project, attempts were made to develop a fish assemblage indicator using the same procedure as for developing a benthic invertebrate assemblage indicator. Each step in that procedure is outlined below.

Step 1: Identify candidate fish assemblage measures

Fourteen fish measures were identified as candidates for discriminating between sites of differing environmental quality (Table 4-9). The list of candidate measures was developed based on general knowledge of fish assemblages and their response to stress, and on historic case studies examining the effects of local perturbations on fish assemblages. Most of the case studies involved freshwater fisheries, particularly those considering application of the IBI (Karr et al. 1985; Berkman et al. 1986; Leonard and Orth 1986). Measures suggested by Miller et al. (1988) for estuarine application of an IBI and measures identified in a number of estuarine case studies investigating the response of fish assemblages to site- specific perturbations (Bechtel and Copeland 1970; White et al. 1977; Haedrich and Haedrich 1974; Horn 1980; Allen et al. 1983) also were incorporated.

Identification of appropriate species composition measures presented the greatest challenge in developing the list of candidates. Species composition can be quantified in two ways. The first is the indicator species or the indicator taxa approach. In this approach, the presence/absence or dominance of a particular taxa is used to define the quality of the site. Presence/absence is appropriate to sensitive or specialized species, such as endangered organisms, whereas dominance might be more appropriate for abundant or ubiquitous species. The second is the assemblage tolerance approach developed by Hilsenhoff (1987) for freshwater benthic invertebrates. For this approach a tolerance value is assigned to each taxon and used to calculate an abundance-weighted tolerance value for the assemblage. This approach differs from the indicator taxa approach in two ways. First, it emphasizes the most abundant taxa at the site; consequently, rare taxa, such as endangered species, that might be accurate indicators of a high quality environment would contribute little to the ranking of the site. Second, it requires ranking all taxa in the assemblage, not just those for which specialized knowledge about their preferences and tolerances is available.

Table 4-9. Tests for significant differences among degraded and reference sites for each candidate fish measure

	t-test (p-value)	Direction (+ = higher value at reference sites)
Measures of Biodiversity		
Shannon-Weiner index	0.04	+
Evenness	0.02	+
Number of species per haul	0.10	+
Measures of Community Condition		
Total number of fish per haul	0.22	+
Measures of Individual Health		
Percent of fish with gross pathology	0.01	-
Measures of Functional Groups		
Percent top carnivores	0.65	-
Percent pelagic invertivores	0.91	-
Percent benthic invertivores	0.04	+
Percent planktivores	0.77	+
Measures of Taxonomic Composition		
Composition index	0.02	+
Percent bullheads	0.82	-
Percent clupeids	0.75	-
Percent scianids	0.44	+

The difficulty with applying these approaches to the 1990 EMAP-E data is that neither has been elaborated for estuaries. Both require information on the relative tolerance of estuarine fish to pollution exposure that is not readily available. Most studies to assess differences in fish assemblages between degraded and nondegraded sites have been limited to site-specific, discharge-related problems, rather than to more generalized habitat degradation problems. Almost all have been conducted on local scales, usually within a single system, and most have been limited to a narrow range of salinities.

Data from the 1990 Demonstration Project was used to develop a composition index, based on the approach used by Hilsenhoff (1987). This approach requires tolerance values ranked from 0 to 10 for each species encountered in the samples. No such values exist for estuarine

fish, so they were generated based on presence/absence information at the test sites (Table 4-3). The tolerance value was set equal to the percent of reference sites at which the species was found minus the percent of degraded sites at which it was found, normalized to a scale from 0 to 10. These procedures were applied separately to three salinity zones (less than 5 ppt, 5-18 ppt, greater than 18 ppt) to account for the possibility that a fish's degree of tolerance for degraded conditions may be a function of its proximity to the extreme boundary of the natural habitat of the species. The composition index for each station was calculated as the abundance-weighted average tolerance value.

Each of the candidate measures was tested for relationships with salinity, depth, and latitude; substrate was not tested because a trawl covering several hundred meters typically encountered a wide range of substrate types. None of the measures were significantly correlated with latitude. Abundance and species richness were the only measures that correlated significantly with salinity. The relationship with salinity was greater for abundance than for species richness, but the r^2 value did not exceed 0.1 for either of these. The relationship with depth was significant only for the composition index, but less than 5% of the variation was explained by the relationship. Because the amount of variance accounted for by environmental factors was small, no adjustments were made to the data before they were used in the discriminant analyses.

Step 2: Develop a test data set

The test sites used to evaluate the fish response were the same as those used to evaluate the benthic response (Table 4-3). The original intent was to use data from sampling interval 2 to calibrate the fish indicators and from sampling interval 3 to validate the consistency of responses; however, because fewer than 40 fish were caught at more than half of the stations in interval 2, data from sampling intervals 2 and 3 had to be combined for fish analyses. Data from interval 1 were not used because of the large number of sites that were not sampled in that interval, and because it had been determined previously, based on analysis of the water quality data, that sampling in future years will be concentrated between late July and early September (see Section 5.1).

Step 3: Identify combinations of candidate measures that discriminate between degraded and nondegraded areas

When subjected to univariate examination, only five measures, prevalence of pathology, evenness (number of species constituting 90% of the assemblage), the Shannon-Weiner diversity index, the composition index, and percent of benthic-feeding fish in the assemblage, distinguished significantly between degraded and nondegraded sites in the indicator testing data set (Table 4-9). Visible pathology was the most effective measure. The prevalence of visible pathology was less than 1% at all reference sites, and one-third of the polluted sites exhibited prevalances higher than 1%.

Three measures, prevalence of pathology, evenness, and the composition index, were identified by stepwise discriminant analysis as the most appropriate combination of measures for discriminating between degraded and reference sites (Table 4-10). When this group of measures was weighted optimally, successful discrimination was achieved for 13 of the 14 (93%) reference sites, and 16 of the 19 (84%) degraded sites. All three degraded sites that were misclassified were located in areas north of the Chesapeake Bay and were classified as degraded on the basis of high concentrations of contaminants. No pattern or trends were observed in the habitat variables among the misclassified sites. When the analysis was repeated after eliminating station 106 from the calibration data set, as in the analysis for benthic invertebrates, the same three candidate measures were selected in discriminant analysis. Eliminating this station affected the parameter coefficients slightly but had no effect on the number or identities of misclassified stations.

Step 4: Validating the model

Two potential fish indices were developed in step 3. The cross-validation efficiency for Index 1, which included station 106, was 85%; for Index 2, which did not include station 106, the cross-validation efficiency was 83%. Although calibration of the fish index was conducted with combined data from intervals 2 and 3, the index was recalculated for each period individually to examine whether the response was consistent at each sampling site. Of the 19 indicator testing sites for which data were available in both intervals, 16 (84%) were classified the same between periods. At two of the three sites that were classified differently, calculation of the index was based on a collection that included less than five fish in at least one of the periods; an even higher performance might have been achieved by eliminating variability associated with small sample sizes.

Validation against an independent data set is particularly important for the fish index because the composition index was included in the final discriminant function. Inclusion of the composition index involves some circular reasoning because the tolerance values on which it is based were derived using the same data set used for examining discrimination

Table 4-10. Results of discriminant analyses conducted for combining candidate fish measures into an index

Analysis	Selected Measures	Calibration Efficiency		
		Percent of Degraded Sites Correctly Classified	Percent of Reference Sites Correctly Classified	Canonical R^2
Index 1 • All variables included	•Percent of fish with gross pathology •Eveness •Composition index	84%	93%	0.49
Index 2 • All variables included • Station 106 removed	•Percent of fish with gross pathology •Eveness •Composition index	83%	93%	0.50
Index 3 • Composition index eliminated	•Percent of fish with gross pathology •Shannon-Weiner index	84%	71%	0.34

between degraded and reference sites. Recognizing this shortcoming, stepwise discriminant analysis was repeated excluding the composition index. Under this scenario, only percent pathology and the Shannon-Weiner index were selected in the analysis (Table 4-9). The rate of discrimination using just these two measures was 79%. The cross-validation efficiency also was 79%. When period 2 and 3 data were compared, however, less than half of the sites were classified the same (degraded or not degraded) between periods using this index. The poor agreement between intervals 2 and 3 using this combination of measures may be due, in part, to small sample size; however, it suggests the importance of the composition index to a stable multi-metric fish index of environmental quality.

Step 5: Scaling the index

None of the potential fish indices developed above were validated sufficiently to warrant placing them on a scale from 0 to 10 or incorporating them into the preliminary evaluation of environmental condition presented in Section 2. Although index 1 appears promising, use of the composition index requires further validation. Index 3, which did not include the composition index, had less than 80% discrimination and compared poorly between periods.

4.1.3.2 Discussion of the Fish Index

Developing an indicator of environmental quality based on fish assemblages may be more difficult than developing one based upon benthic invertebrates for several reasons. First, the mobility of fish makes them more difficult to collect; the efficiency of trawls is typically low and species-specific (Kjelson and Colby 1977). Second, fish assemblages in an estuary are often transient; thus, the assemblage at any one time may not accurately reflect conditions at a site. Third, fish respond to habitat variables (e.g. structure) that are difficult to measure or incorporate in an index.

To assess whether the trawl sampling gear used in the Demonstration Project provided a consistent description of the fish assemblage at a site, correlations were identified for each measure between the first and second hauls at a site in interval 2 (replicates were collected at a subset of sites in interval 2). The correlation was significant for every measure, and the r^2 for most measures exceeded 0.75 (Table 4-11). Presumably the large gear combined with efforts to standardize trawls resulted in this consistency.

To assess whether the assemblage at a site was transient over the sampling period, correlations were conducted for each measure between the first trawls taken at a site in intervals 2 and 3. Correlations for the biodiversity measures, percent pathology, and the composition index were significant, although none had an r^2 as high as 0.5 (Table 4-11). The correlations for measures relating to taxonomic composition and functional groups were not significant.

Table 4-11. Correlations between the first and second trawl and between intervals 2 and 3 for each candidate fish measure

	Trawl 1 vs. Trawl 2	Sampling Interval 2 vs. Sampling Interval 3
Measures of Biodiversity		
Shannon-Weiner Index	0.65	0.40
Evenness	0.74	0.64
Number of Species	0.87	0.70
Measures of Community Condition		
Total Number of Fish per Haul	0.84	0.37
Measures of Individual Health		
Percent of Fish with Gross Pathology	0.89	0.87
Measures of Functional Groups		
Percent Top Carnivores	0.63	0.03
Percent Pelagic Invertivores	0.86	0.38
Percent Benthic Invertivores	0.72	-0.12
Percent Planktivores	0.82	-0.02
Measures of Taxonomic Composition		
Composition Index	0.81	0.62
Percent Bullheads	0.95	-0.05
Percent Clupeids	0.80	-0.04
Percent Seianids	0.93	0.18

The issue of transience is exacerbated by the fact that most fish populations respond to environmental impacts on a cumulative, basinwide or regional level. Frequently, estuarine fish are part of coastal stocks and complete their life cycles in portions of different estuaries; consequently, the abundance and composition of fish may not respond on the same spatial scale as the Demonstration Project sampling activities. While this would not eliminate fish as useful indicators, it does present a difficulty for validating them, particularly given the site-specific orientation of the validation approach used here.

Failure to account for all habitat variables at a site probably also hinders development of a fish index that discriminates degraded sites from nondegraded reference sites. At each of the sampling sites, habitat was characterized by two variables, salinity and depth. Fish abundance is related to many additional habitat variables, such as bottom type, presence of

structure, and distance from shore. Some measures will not distinguish well without normalizing for these additional habitat variables. For example, some fish species are attracted to submerged structures for protection and food resources. Several of the degraded sites in the indicator testing data set were located in small systems with substantial physical structure (e.g. docks, debris such as cars, which serve as artificial reefs), whereas some of the reference sites were located miles from shore in areas with a uniform bottom. Accounting for these differences is difficult but may be required to develop a successful fish index.

Some of these difficulties could be reduced by conducting shore-zone sampling with a seine net instead of offshore sampling with a trawl. Shore-zone sampling dramatically reduces the amount of habitat variability because seine sampling is limited to low gradient beaches with little structure. Additionally, the shore-zone assemblage typically includes smaller, less mobile fish and is more stable over the summer period. Incorporation of shore-zone sampling into EMAP, however, would be difficult. The percentage of shore- zone area that is sampleable is limited, and sampling it in an unbiased manner would require a different sampling design than is being used in the rest of the program.

Despite the apparent difficulties associated with developing a fish index based on trawl sampling, the results of index 1 showed some promise for discriminating environmentally degraded areas on a regional scale. Further testing, validation, and refinement should be pursued. Three areas appear to be most appropriate for future effort: 1) refining tolerance values used in the composition index, 2) developing a data set for validating the combination of measures identified during the 1990 Demonstration Project, and 3) identifying new measures that could be used in addition to those already identified. Several fish measures, such as tissue contaminant levels and biomarkers, have not yet been incorporated into the index development process.

4.1.3.3 Visible Pathology

Regardless of whether a fish index that successfully discriminates sites of high and low quality can ever be validated, there are still a number of fish measures that convey meaningful information for describing trends on a watershed basis. These measures include, the percent of fish with visible pathological disorders, the percent of fish of commercial/recreational value, the percent of introduced species, or the percent of estuarine area (or range) over which a particular species of interest is found. These types of indicators cannot be used effectively in some types of status evaluations because a threshold boundary for acceptable and degraded condition cannot be defined easily; however, these indicators are useful for comparing status among watersheds. They would be particularly useful for assessing trends in watershed or regional condition because a change in indicators, such as an increase in the range or spatial coverage of a nuisance fish (e.g. lampreys), or a decrease in the percent of fish with visible pathological disorders, can be interpreted relatively easily.

Estimates of the prevalence of visible pathological disorders provide a good example of how data for which baseline conditions are not known can still be useful for assessment. It was possible to demonstrate large differences in the prevalence of visible pathology among different classes of estuaries using 1990 Demonstration Project Data (Section 2). Even without knowing whether a specific rate of pathology is considered above background levels, it was possible to identify that fish in small estuarine systems had considerably more pathological problems than fish in large estuaries or large tidal rivers. This is important information because the frequency of pathological disorders in fish is often used as an indicator of environmental quality not only by scientists, but also by the general public. The public associates fish pathologies with poor water quality. An increase in visible pathologies on recreationally and commercially valuable species may result in adverse public reaction and subsequent economic consequences. The relationship between environmental stress and visible pathological conditions such as fin erosion, skin tumors, and ulcers in fish has been established, and highly polluted habitats have greater frequencies of fish with visible pathological disorders than similar, less polluted habitats (Sindermann 1979; Mearns and Sherwood 1974; O'Connor et al. 1987; Buhler and Williams 1988; Malins et al. 1984, 1988). In addition, measuring visible pathological disorders is cost effective because fish can be screened rapidly for abnormalities while being collected for other analyses, such as tissue chemistry or assemblage information.

Although promising, the visible pathology indicator requires further development. Many types of visible pathology like those encountered during the Demonstration Project are species- and size-specific (Moller 1984; Wolthaus 1984); thus, status estimates that compare classes, or trends analyses that compare years, may confound differences or changes in species or size composition with differences or changes in environmental factors that lead to visible pathological abnormalities. EMAP presently is developing normalization procedures that will account for size and species effects on visible pathology in the assessment process.

4.1.3.4 Fish Tissue Contaminants

Several fish measures, such as tissue contaminant levels or biomarkers, have not yet been measured in EMAP-E (tissue samples were collected, but not processed in 1990; see Section 4). The principal problem with using indicators at the level of individual fish is that making regional estimates of environmental condition requires collecting fish from a small target group of taxa at a large percentage of the sampling sites. The target group must be small to minimize difficulties with species-specific response. Since the EMAP design is probability-based, sites where fish from the target group can not be collected present an inference problem: it is not possible to determine whether failure to capture target species resulted from insufficient collection effort or from adverse (e.g. contaminated) conditions that caused fish to avoid those sites. If the latter is true, failure to include those sites in the evaluation would

result in underestimating the extent of estuarine area with subnominal environmental conditions.

Data collected during the 1990 Demonstration Project were used to determine whether a small number of target species could be captured at a sufficient number of sample sites to justify generating estimates of condition for the Virginian Province. Ten target species were identified as the most appropriate to provide provincewide coverage, based on a survey of historic fish trawling information (Holland 1990). These species were retained for tissue chemistry analysis during the Demonstration Project (Table 3-5). At least one of the target species was captured at 75% of the sites sampled during the 1990 Demonstration Project.

Most of the sites where target species were not collected were too deep to trawl safely using the hydraulic winch system available on sampling boats. Underwater obstructions prevented successful trawling at less than 6% of the sites. With minor methodological modifications and slight alterations in the list of target species, it is anticipated that target species will be collected at more than 90% of the sites in future years. On this basis, collections for contaminant analysis will be continued in future years of the program, and biomarker research within the program will be expanded.

4.2 EXPOSURE INDICATORS

Methods for measuring exposure indicators at specific sites are well-developed and have been used in a number of monitoring programs throughout the province. Methods for measuring indicators on a regional scale in a probability-based manner to develop regional status estimates of the kind being made in EMAP are not well-developed.

4.2.1 Dissolved Oxygen

Dissolved oxygen (DO) is a fundamental requirement for the maintenance of fish, shellfish, and other aquatic biota. Dissolved oxygen concentrations also reflect the integrated response of complex natural systems to nutrient input. The increased prevalence of low dissolved oxygen conditions in geographically dispersed estuaries is a central aspect of the declining health of near coastal waters (Turner and Allen 1972; May 1973; Jorgensen 1980; Harper et al. 1981; Rossignol-Strick 1985; Justic et al. 1987; Kuo and Neilson 1987; Rosenberg 1990). Regulatory programs to improve dissolved oxygen conditions in estuaries have cost billions of dollars, and assessing the regional effectiveness of these programs is one of the goals of EMAP.

The major problem with using dissolved oxygen as an exposure indicator is its extreme temporal variability in estuarine waters (Breitburg 1990; Sanford et al. 1990). Dissolved

oxygen concentrations in bottom waters can fluctuate from supersaturation to severe hypoxia within hours. Factors contributing to these fluctuations include tides, meteorological conditions, and biological activity. In stratified estuaries, bottom waters move with tidal currents and seiches, and if partially depleted of oxygen, bottom water DO measurements will fluctuate in response to water movements. During the 1990 Demonstration Project, continuously-recording DO meters were deployed at sites throughout the Virginian Province to document the extent of these fluctuations in bottom waters. Examples of DO fluctuations in the Virginian Province are shown in Figure 4-4.

The purpose of this section is to use continuously-recorded DO data collected during 1990 to identify the most appropriate method for measuring DO within EMAP's sampling constraints. These constraints included safety considerations, time and funding limitations, and equipment performance limitations. For safety and logistical reasons, stations were sampled during daylight hours only. Stations could be visited a maximum of three times during the summer index period due to limited time and funds. Dissolved oxygen meters could be deployed reliably for no more than three days at a time. The duration of deployment was limited because fouling reduces the reliability of measurements after about three days. Deployments beyond three days also exceeded the funding and logistical limitations of EMAP.

Given these constraints, three sampling strategies were evaluated:

- Single point-in-time samples collected on random days in the index period between 0800 and 1800 hours.

- Multiple point-in-time samples collected on random days in the index period, visiting each site two or three times. The index period is divided into two or three equal intervals, and sample dates are chosen randomly from each sub- period. Successive point samples at the same station are separated by at least nine days to avoid autocorrelation problems.

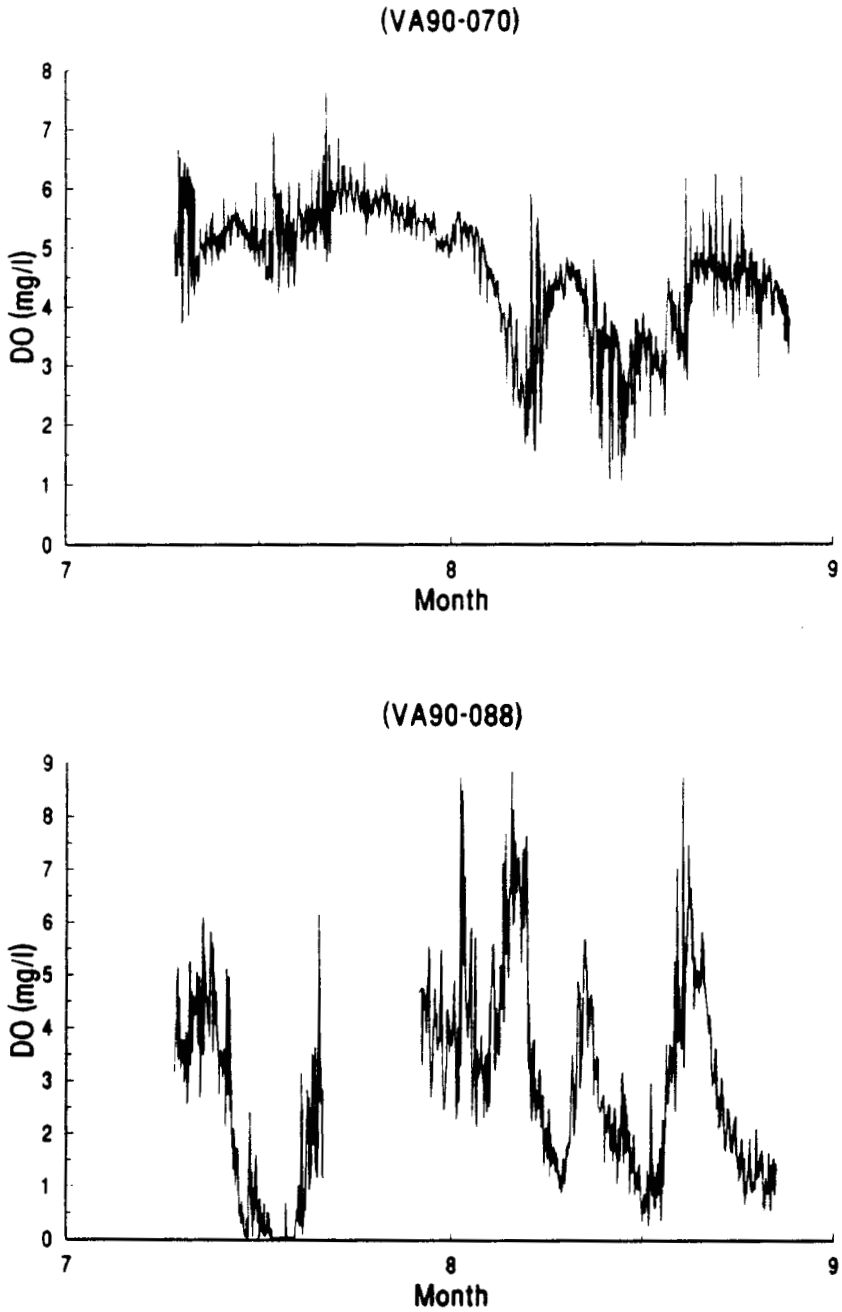

Figure 4-4. Time series data obtained at two long-term dissolved oxygen monitoring sites showing fluctuation of dissolved oxygen concentrations over the summer index period.

- Short-term continuous records obtained by deploying a continuously-recording meter at each site to measure DO at 30-minute intervals for one, two, or three days. The starting date for the three-day period is selected randomly within the index period.

Making three-day continuous measurements was the most capital-intensive of the three alternatives, as it required purchasing a sufficient number of units for simultaneous deployment at several stations, and extra units for calibration, repair, and replacement of lost units. Three-day deployments, however, are not as labor intensive as three single-point measurements and are approximately equal in labor to collecting two point-in-time samples per station. Lost data may be a problem because of lost or malfunctioning meters, but 1990 data suggested that such losses would occur less than 3% of the time (see Section 3); furthermore, point measurements (CTD casts) taken upon meter deployment and retrieval serve as a backup estimate if the continuously-recording meter fails.

Dissolved oxygen data were used in two ways: 1) for estimating status and trends, in which the objective was to evaluate the condition of estuarine areas rather than individual sites; and 2) for identifying associations, which required a higher degree of certainty about conditions at an individual site. Measurement methods were evaluated separately for each use.

4.2.1.1 Use of dissolved oxygen data for estimating status and trends

Developing a regional status estimate involves generating a population distribution estimate (a cumulative distribution function, or CDF) from the sample data. In EMAP-Estuaries, samples were collected over an index period lasting 30 to 60 days, and sample sites were visited in a semi-random fashion (complete randomness is not possible due to logistical constraints). The CDFs used to make EMAP-E status estimates were calculated from the sample data and site-specific inclusion probabilities.

Point-in-time Measurements

Point-in-time measurements represent the least expensive and most logistically feasible means of estimating the status of dissolved oxygen exposure. Given that it is not possible to sample an entire province synoptically, the CDF developed from point-in-time samples depended on the order and time period in which the samples were taken. This will especially be true if DO fluctuates systematically between sites, resulting in an unstable distribution of DO. Using single, point-in-time measurements to make status estimates requires stability in the areal extent of low DO over the province throughout the index period.

The stability of point-in-time measurements was tested by comparing the CDFs of bottom dissolved oxygen concentrations between two successive sampling intervals (Fig. 5-5). These comparisons demonstrated that the distribution of dissolved oxygen was stable between the two periods; thus, single point-in-time measurements of dissolved oxygen are adequate for estimates of status on a regional basis during the summer index period.

Multiple Point-in-time Measurements

Multiple point-in-time measurements changed neither estimates of status, nor the shape of the distribution; however, the increased sample size of multiple point-in-time measurements reduced sample variance and increased the power to estimate trends.

Continuous Oxygen Records

Continuous samples of one to three days duration were marginally better for obtaining status estimates. The primary advantage of the short-term continuous samples was the ability to remove the sampling bias due to collection of only daytime samples by allowing selection of a single random sample from each continuous record. This results in the same sample size as single point samples. The additional data points obtained by continuous samplers were of little use in estimating status because the DO time series is strongly autocorrelated, and samples within three days of each other are not independent (Fig. 4-6).

The effects of natural, daily fluctuations in DO concentration must be considered. EMAP sampling was carried out during daylight hours for logistical and safety reasons, and stations typically were sampled between the hours of 0800 and 1800. Water column dissolved oxygen concentration is determined by photosynthesis, respiration, and atmospheric diffusion. In productive shallow waters, there may be a circadian cycle of DO in which a minimum concentration occurs around sunrise, and a maximum concentration occurs near sunset. Diurnal periodicities in DO also occur. They were most pronounced in waters less than 3 m deep, while tidal periodicities are most pronounced in waters more than 6 m deep (Fig. 4-7). Deep waters are more likely to be stratified,and subpycnocline hypoxic water can move in and out with the tide, producing a tidal signal in the DO record. The daytime sampling schedule of EMAP may have resulted in biased estimates of dissolved oxygen exposure.

We tested the difference between daytime and completely random sampling, and found that the advantage of complete temporal randomization is virtually undetectable, based on Monte Carlo sampling of the long term DO (LTDO) time series. One hundred random observations were taken from each of 13 time series, and 100 random observations were

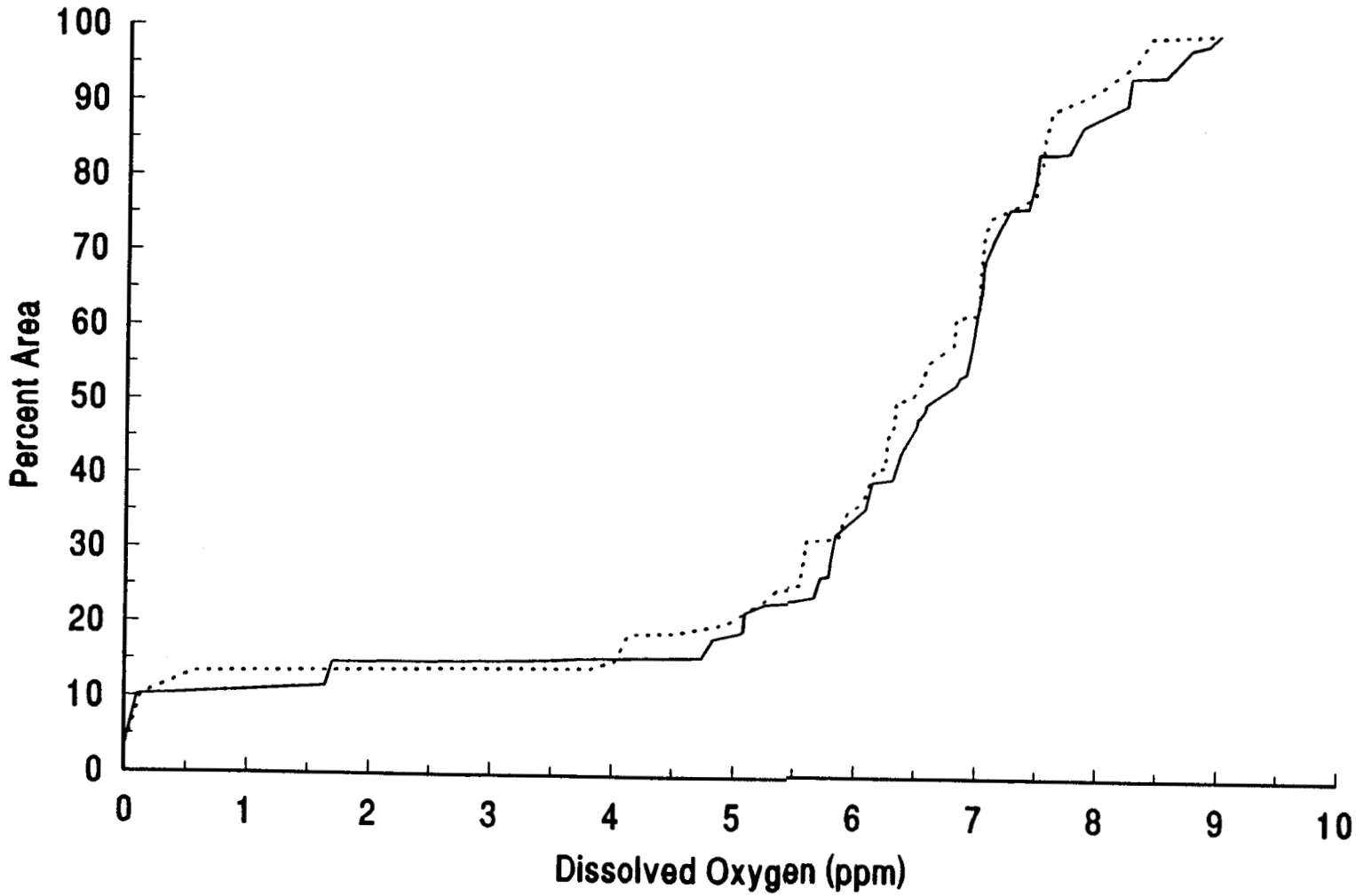

Figure 4-5. Comparison of CDFs for province-wide bottom dissolved oxygen concentrations in sampling intervals 2 (solid line) and 3 (broken line)

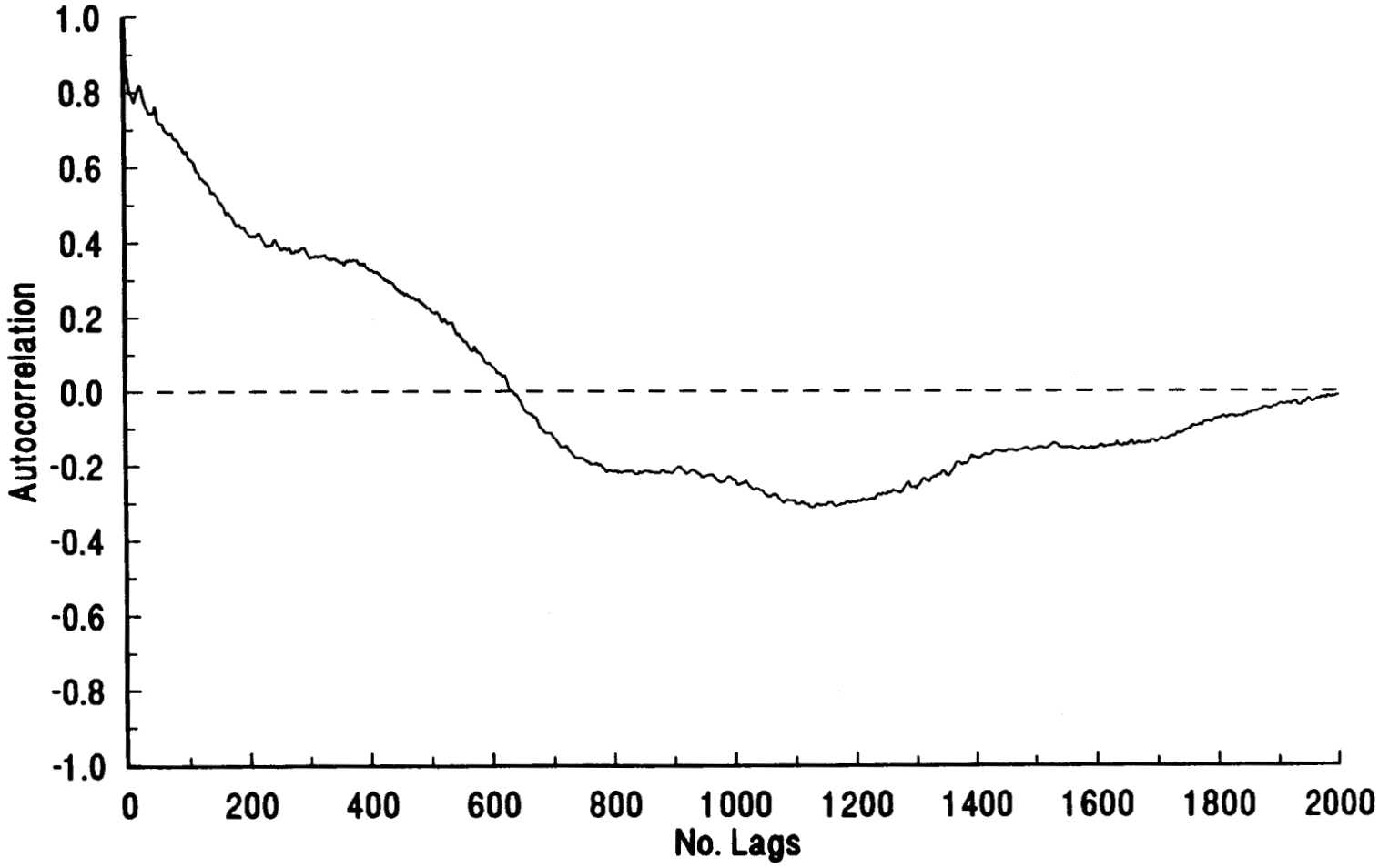

Figure 4-6. Example showing the degree of autocorrelation on samples from continuous monitoring of dissolved oxygen concentration. Station from which these data were taken is located in Narragansett Bay (1 lag = 30 minutes).

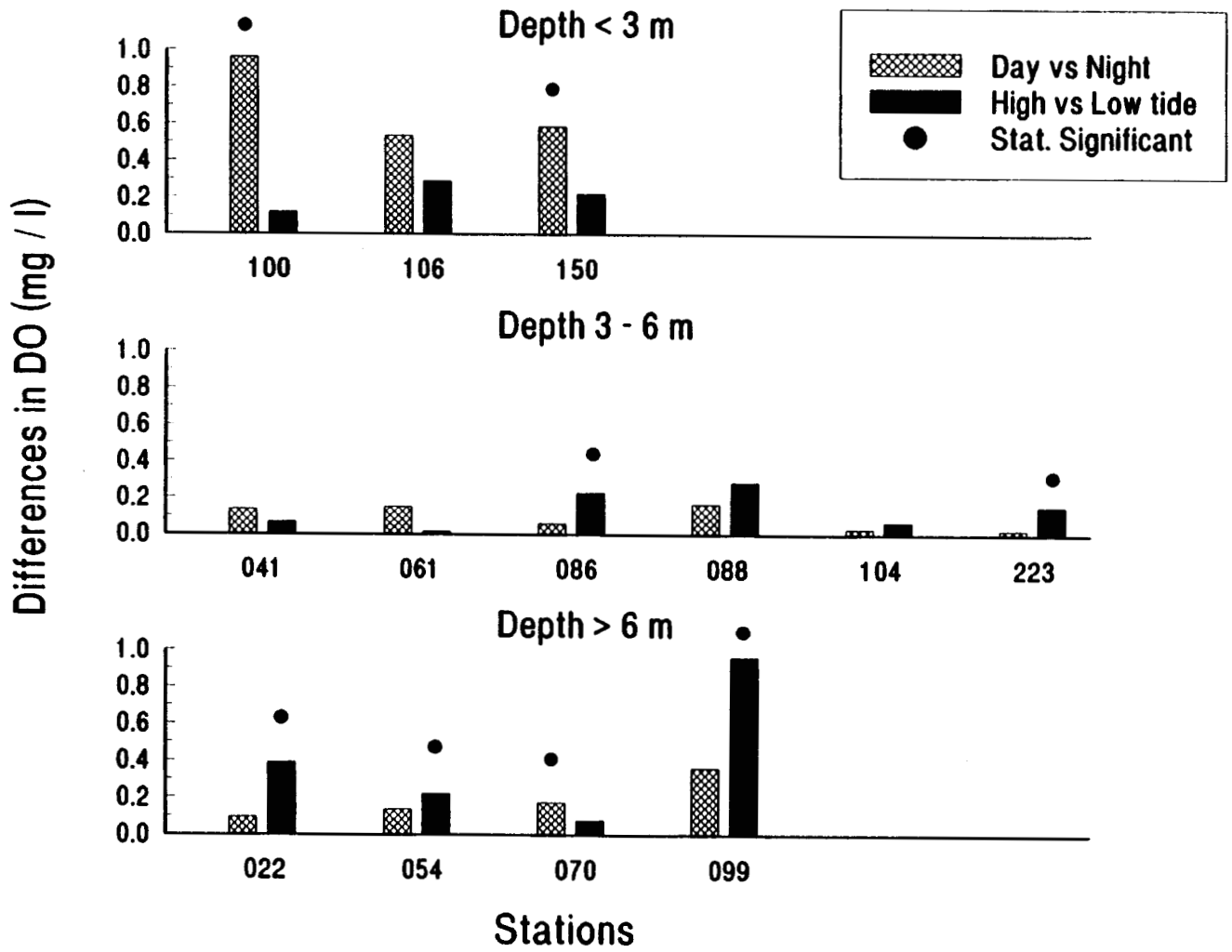

Figure 4-7. Differences in dissolved oxygen due to diurnal and tidal periodicities at stations of different depths.

taken from observations between 0800 and 1800 hr in each time series. Samples from the time series were combined to develop a CDF for each sampling strategy (Fig. 4-8). The two sample CDFs were not significantly different from each other, and both adequately represented the population CDF.

Only a few of the LTDO sites displayed a relatively strong diurnal signal (Fig. 4-7). On a site basis these patterns are very important. The Monte Carlo test suggested that on a regional basis, the effect of those sites may be unimportant; however, this conclusion is dependent on the representativeness of the LTDO sites. The importance of these cycles, hence, the importance of randomizing with respect to them, depends on the relative frequency of sites where diurnal cycles are dominant.

4.2.1.2 Use of the dissolved oxygen indicators for associations

The objective of examining associations is to identify relationships between ecological response and environmental exposure. Although this can be done by comparing the coincidence in trends of exposure and response in a population, the most powerful methods require paired observations of DO and response indicators at each site, allowing regression or similar analytical approaches.

There are two substantive problems in identifying associations with DO. One is that it is not clear which DO parameters (e.g., mean, minimum, % of time below a critical value, duration below critical value) are critical for survival of organisms and for controlling the responses of ecological communities. Although most estuarine organisms can tolerate short exposures to dissolved oxygen concentrations below saturation without apparent adverse effects, prolonged exposures to less than 60% oxygen saturation (~ 5 ppm) may result in altered behavior, reduced growth, reduced reproductive success, and mortality (Kramer 1987; Burton et al. 1980; Reish and Barnard 1960). Exposure to less than 30% saturation (~ 2 ppm) for one to four days is lethal to most biota, especially during summer months, when metabolic rates are high. Stresses that occur in conjunction with low dissolved oxygen (e.g., exposure to hydrogen sulfide or ammonia) may cause more harm to aquatic biota than exposure to low dissolved oxygen concentrations alone (Brongersma- Sanders 1957; Adelman and Smith 1972; Theede 1973). Finally, aquatic populations exposed to low dissolved oxygen may be more susceptible to the adverse effects of other, unrelated stresses.

The second problem relates to our ability to measure DO parameters. Obviously, duration or percent of time below a critical value cannot be estimated with a single point-in-time measurement, or even a few such measurements. If the measured DO does not reflect "true" exposure to low DO, then EMAP's ability to identify associations between low DO and subnominal values of response indicators will be limited. As shown in Figure 4-4,

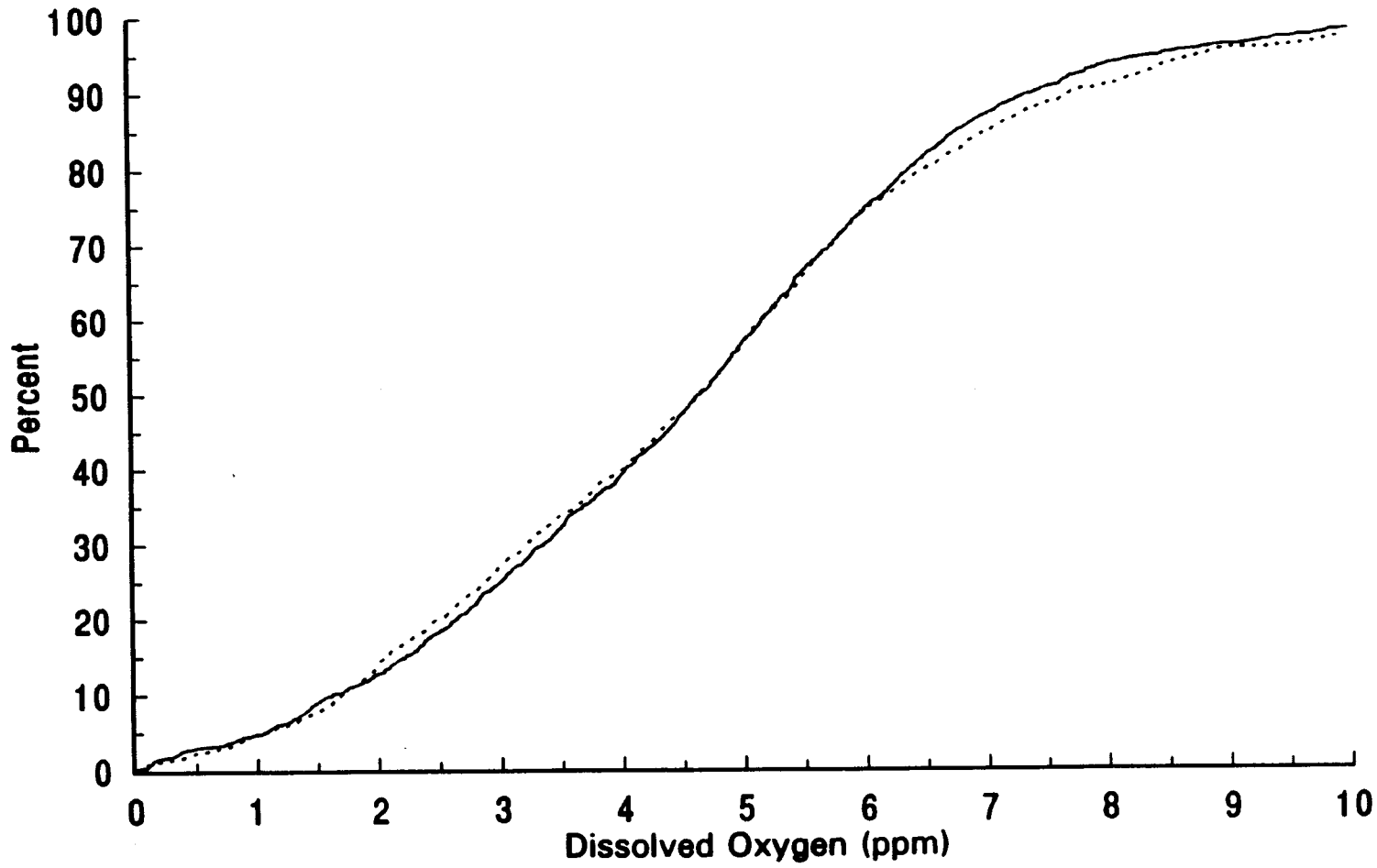

Figure 4-8. CDFs developed by restricting sampling of time series data to daylight hours (broken line) and by unrestricted random sampling (solid line)

bottom water dissolved oxygen concentrations can fluctuate between near saturation and severe hypoxia in a matter of days (also see Breitburg 1990; Sanford et al. 1990). A single moderate or high DO measurement may not represent hypoxia that occurred before or after the sample was taken.

During the Demonstration Project, three sampling strategies were examined for their ability to characterize DO at a site with sufficient precision for investigating associations. Once again, Monte Carlo sampling of the LTDO data sets was used to simulate realistic methods of sampling within EMAP constraints. Sampling strategies were one, two, and three point-in-time samples and one, two, and three-day continuous sampling. For two and three point-in-time samples, a random daytime (0800-1800 hours) sample was taken so that samples were separated by at least nine days to avoid autocorrelation problems. Estimators of the mean DO at a site were examined, as well as estimators that classified a site as above or below a threshold DO value (e.g., 25% of the time above 2 ppm). It is not necessary to develop good estimates of a parameter to identify associations; determining whether the parameter exceeds a threshold is sufficient to produce a categorical variable, allowing categorical data analysis.

Performance of a sampling strategy in estimating the mean of a time series was evaluated by determining the frequency of samples that were more than 1 ppm away from the true site mean. Single point sampling performed the worst for estimating the site mean (52% error rate above 1 ppm (Fig. 5-9). The one to three day continuous samples were not much better (39% to 44% error rate). Three single points performed best for estimating the site mean, with a 22% error rate, followed by two single points at a 28% error rate. The reason for the poor performance of the short term continuous samples is clearly the strong autocorrelation of dissolved oxygen within a three-day sample period. A three-day continuous sample is slightly better at estimating the site mean than a single sample, but not as good as two independent samples.

The threshold used in the analysis was 25% of time below 2 ppm. Three of the 13 LTDO sites were below this threshold (Fig. 4-10). Two measures were investigated, the mean and the minimum of the observations. Alternative sampling strategies were tested by determining their ability to discriminate between LTDO data sets above and below the threshold, and calculating the misclassification rate for each threshold value for each sampling alternative (Fig. 4-11). The overall misclassification rate for each strategy was the rate at which false positive (labeling a good site as bad) and false negative (labeling a bad site as good) classifications were the same.

Estimators of threshold exceedance performed far better than estimates of the site mean. The single point-in-time sample was, again, the poorest, with a 26% overall misclassification rate (Fig. 4-12). Measures based on the minimum of two or more point-in-time observations performed better. The minimum of three point-in-time samples was the best

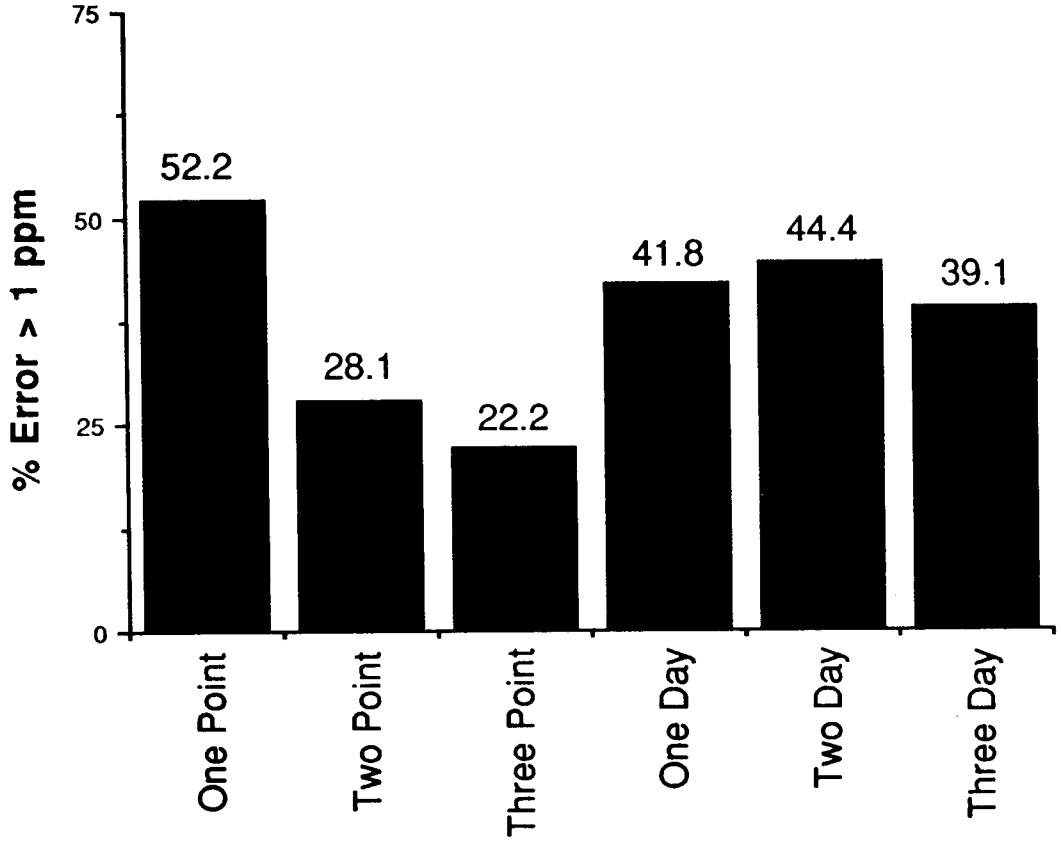

Figure 4-9. Relative performance of the six sampling strategies in estimating the mean dissolved oxygen concentration within 1 ppm from time series

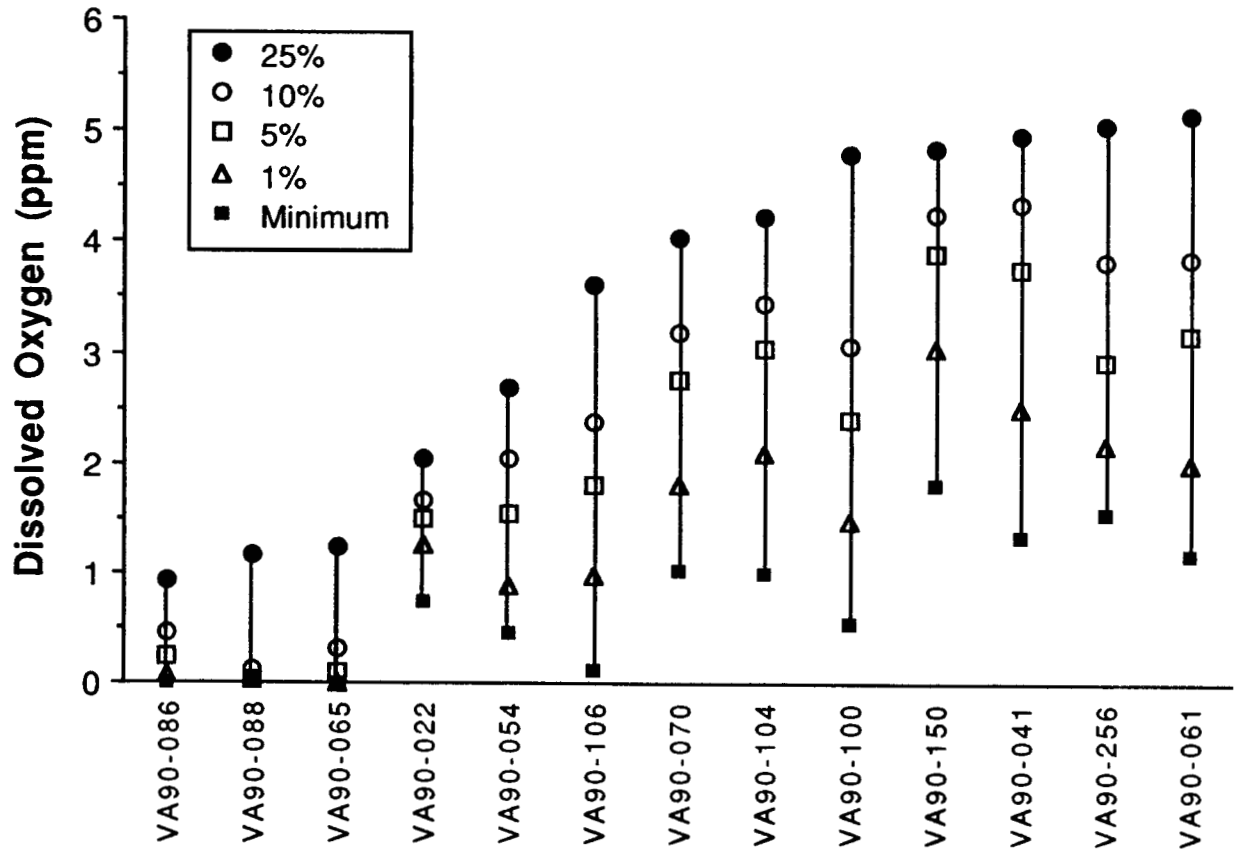

Figure 4-10. Dissolved oxygen quantities from minimum to 25%, for 13 LTDO data series.

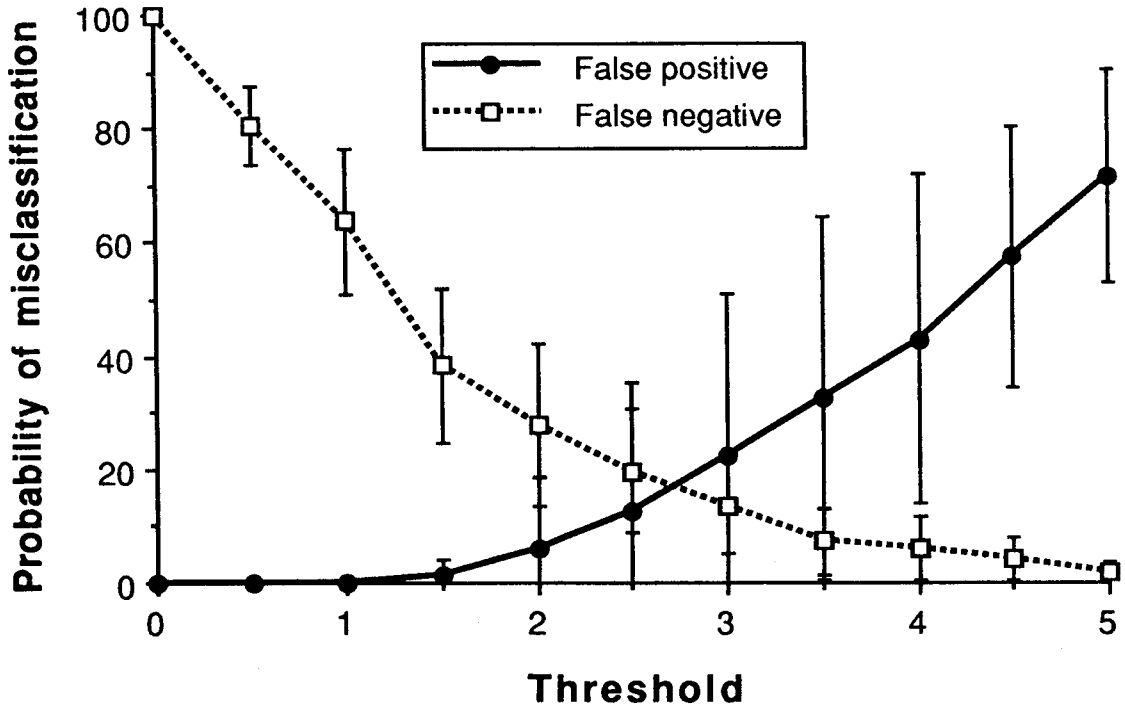

Figure 4-11. Misclassification rate associated with DO threshold values for one sampling strategy. The point where the lines cross is the overall misclassification rate (where rates of false positive and false negative classifications are the same.)

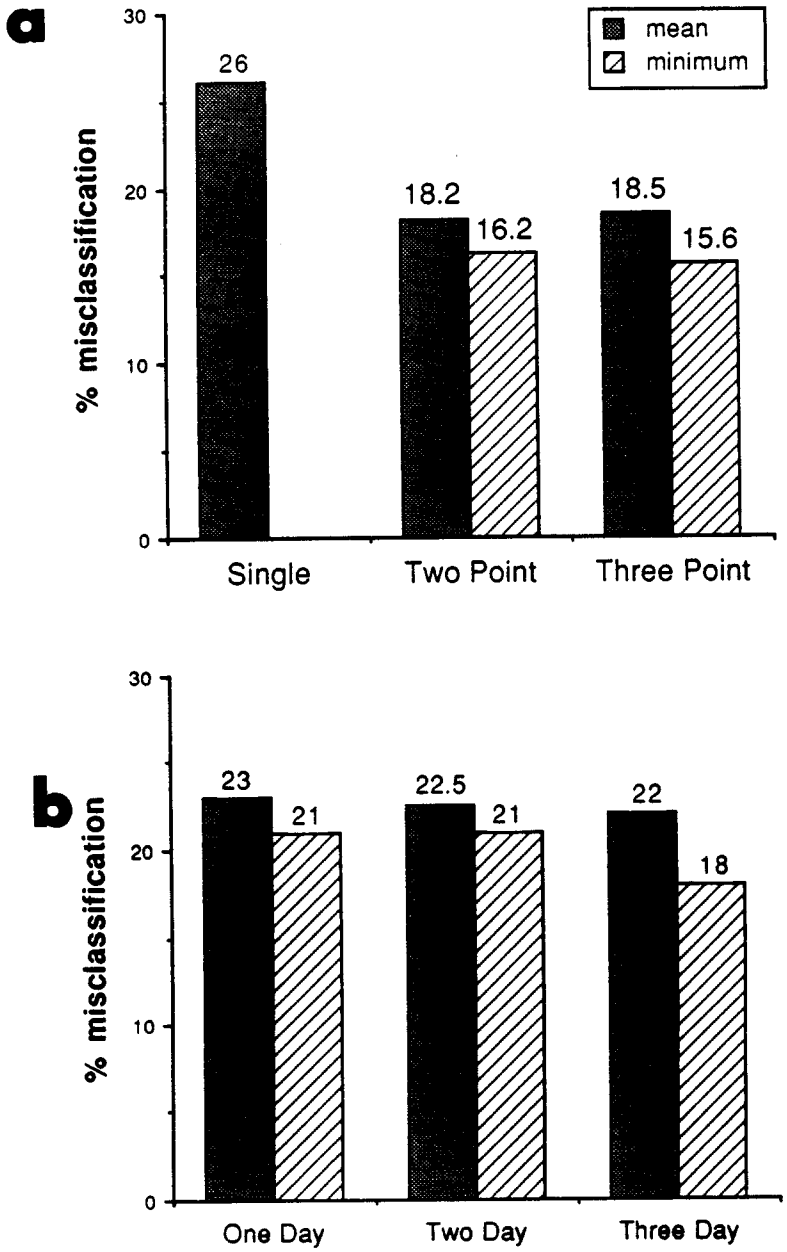

Figure 4-12. Overall misclassification rates for point-in-time sampling (a) and short-term continuous sampling (b) derived from plots like the one shown in in Figure 4-13.

overall, with a misclassification rate of 15.6%. Two point samples also performed well, misclassifying 16.2% of the time, which was slightly better than three-day continuous samples, with 18% misclassification.

Multiple point-in-time samples performed better than short-term continuous samples because of the autocorrelation problem in the continuous samples. It is crucial that the single samples are separated by a minimum time period (in the Monte Carlo test case nine days minimum separation was used) and that they are randomized. Sampling on a fixed interval, even as long as nine or ten days, reintroduces the autocorrelation problem.

Optimum sampling of a serially autocorrelated process depends strongly on periodicities in the process. Sampling should attempt to capture a large part of the variance of the process; if most of the variance is expressed in a short-term diurnal cycle, then short-term continuous sampling will be optimal. Dissolved oxygen at many sites in the Virginian Province appears to have periodicities longer than several days; therefore, two or more point-in-time samples performed slightly better for characterizing those sites than did short-term continuous samples.

The LTDO data sets used for the Monte Carlo sampling are not a random sample from the Virginian Province. Few of the sample time series were characterized by diurnal periods, with the result that multiple samples performed slightly better than short-term continuous samples. A greater predominance of diurnal periods would make short-term continuous sampling the better strategy, especially for accurate estimation of status. Three-day continuous samples are being taken at all sites in the Virginian Province during 1991 and will be evaluated to assess the importance of diurnal periodicity. The dissolved oxygen sampling strategy for the Virginian Province will be refined following the 1991 sampling based on an analysis of these data.

4.2.2 Sediment Toxicity

Sediment toxicity tests are the most direct measure available for estimating the potential for contaminant-induced effects in benthic communities. These tests provide information that is independent of chemical characterizations and ecological surveys (Chapman 1988). They improve upon direct measures of contaminants because many chemicals are bound tightly to sediment particles or are complexed chemically, making them biologically unavailable (USEPA 1989). Sediment toxicity tests have many applications in both marine and freshwater environments (Swartz 1987; Chapman 1988). They have become an integral part of many benthic assessment programs (Swartz 1989), where they are used to establish contaminant-specific effects. The solid-phase, amphipod, 10-day acute test has been employed in sediment assessments in Puget Sound (Dinnel 1990), San Francisco Bay (Long et al. 1990), and New York Harbor (Scott et al. 1990). It is being used currently to examine sediment toxicity in three NOAA National Status and Trends studies in the Hudson/Raritan Estuary, Long Island

Sound, and Tampa Bay (D. Wolfe, NOAA- Rockville, MD, pers. comm.). The success with which this test identifies sediments that are toxic to indigenous biota led to its adoption in the joint Environmental Protection Agency/Corps of Engineers guidance for dredged material permitting (USEPA/ACE 1991).

Although amphipod toxicity test methods have gained general acceptance, a number of factors that affect their application over the broad geographic and habitat range being assessed by EMAP-E. The effects of sediment grain size must be considered. Most coastal contamination is found in fine-grained, high organic carbon sediments, and certain test species have been found to be sensitive to fine-grained sediments (Spies 1989). Similarly, salinity effects are a potential concern. Sediment exposure tests for EMAP were all conducted at the same salinity using full strength seawater (30 ppt salt), regardless of the salinity at the collection site; thus, low salinity sediments were adjusted to full strength salinity for the test. This increase in salinity may have decreased bioavailability of metal contaminants. In addition to the potential effects of habitat parameters such as sediment grain size and salinity, potential effects due to different holding times are also of concern. In a program of this size, processing samples may take as long as 30 days from the time of collection. The effects of holding time on results of sediment toxicity tests, however, have not been well-established (Becker and Ginn 1990).

Chi-square tests were conducted to test the null hypothesis that the relationship between toxicity and contaminant concentration is consistent over a range of salinity, grain size, and holding-time categories in the 1990 Demonstration Project data. For this test, the observed distributions of stations over four conditions (i.e., high sediment toxicity-high contaminant concentration, high sediment toxicity-low contaminant concentration, low sediment toxicity-high contaminant concentration, and low sediment toxicity-low contaminant concentration) for each salinity, grain size, and holding time category were compared to the distribution for the entire province calculated as a percentage of the cases for each category (Table 4-12). Four salinity categories were evaluated: less than 0.5, 0.5 to 5, 5 to 18, and greater than 18 ppt. Three grain size classes were evaluated: less than 20%, 20% to 80%, and greater than 80% silt-clay. Holding time was evaluated for three categories: less than 14, 15 to 28, and greater than 28 days. Stations were characterized as having high sediment toxicity if survival was less than 80% of that in the control. Stations were classified as having high contaminant levels if the concentration of any one contaminant exceeded the median value for biological effects (ER-M) given in Long and Morgan (1990).

Significant differences in the toxicity and contaminant distributions were detected for two conditions: less than 20% silt-clay, and greater than 80% silt-clay (Table 4-12). The differences in distribution between the grain size categories reflect more cases of low sediment toxicity and low contaminant concentrations when the sediments are coarse, and low toxicity with high concentrations of contaminant when the sediments are fine. The latter indicates that high concentrations of contaminants are less toxic in fine sediments. These

results are not surprising because contaminant concentrations are known to correlate well with particle size (i.e., binding capacity is inversely related to particle size).

Table 4-12. Relationship of contaminant concentration to Long and Morgan (1990) ER-M values and toxicity by site, as a function of salinity, grain size, and holding time

Condition	(N)	High Toxicity/ High Contami-nants	High Toxicity/ Low Contami-nants	Low Toxicity/ High Contami-nants	Low Toxicity/ Low Contami-nants	Chi Square
Province	(141)	10	7	15	109	
Salinity (ppt)						
0-0.5	(17)	1	0	3	13	1.62
0.6-5	(16)	1	0	5	10	7.55
6-18	(29)	4	0	3	22	2.78
> 18	(69)	4	6	3	56	5.60
Grain Size (% silt-clay)						
< 20	(57)	1	2	49	49	7.93*
21-80	(22)	5	4	32	32	2.75
> 80	(55)	4	1	21	21	12.80*
Holding Time (days)						
< 14	(18)	1	2	2	13	0.61
15-28	(78)	6	4	5	63	1.63
> 28	(44)	1	0	7	36	2.28

* Statistically different from expected at $p < 0.05$.

Additional tests were conducted as part of the 1990 Demonstration Project to evaluate whether conducting the bioassays at a constant salinity affected the sensitivity of the test. Thirty-eight sediment samples from freshwater/oligohaline habitats were tested with both *Ampelisca* and *Hyalella azteca*, an amphipod used frequently in freshwater sediment bioassays (ASTM 1991). The tests with *Hyalella* were conducted using well water, whereas the tests with *Ampelisca* were conducted using 30 ppt water. In only one of these sediment samples was *Hyalella* survival significantly less than that in the control. For this same set of sediment samples, significant toxicity to *Ampelisca* was found in eight samples. The original concern was that the addition of salt to the low salinity sediments might cause inorganic

contaminants to become less bioavailable, which would have caused the *Hyalella* test to be more sensitive to toxic sediments than the *Ampelisca* test. Instead, the *Ampelisca* test was found to be the more sensitive of the two, suggesting that the salinity adjustment should be of limited concern; however, given the limited number of samples where toxicity to either species was found in this comparison test, additional testing of *Ampelisca* at low salinities, possibly in comparison with other freshwater species, is still warranted.

4.2.3 Sediment Contaminants

The presence of contaminants in estuaries has been identified in both the scientific and popular press as a major problem contributing to degraded ecological resources and restricted harvest of fish and shellfish resources due to human health concerns (Broutman and Leonard 1988; NOAA 1990; OTA 1987; O'Connor 1990). Reducing contaminant inputs and concentrations, therefore, is often a major focus of regulatory programs for estuaries. Contaminants include both inorganic (metals primarily) and organic forms originating from many sources, including atmospheric deposition, freshwater inputs, land runoff, and point sources. These sources are poorly characterized, except in the most well-studied estuaries. Most contaminants that are potentially toxic to indigenous biological resources tend to bind to particles, which ultimately are deposited at the bottom of estuaries (Santschi et al. 1980; Santschi 1984). This binding changes the form of contaminants and removes them from the water column; consequently, contaminants accumulate in estuarine sediments (Santschi et al. 1984; Nixon et al. 1986).

EMAP monitoring efforts have focused on sediment contaminants and have not included measurement of contaminants in the water column (Holland 1990). Concentrations of contaminants in sediments are less variable than those in the water column, and the sediment integrates contaminant inputs to estuaries over months and years. The greater variability of water column contaminant concentrations is due to interactions between contaminant inputs, which are themselves variable, and natural processes in the estuary. Measurements of water column contaminant concentrations made during the EMAP sampling period would characterize estuarine contaminant concentrations poorly and, therefore, are inappropriate.

Sediment contaminant concentrations were measured to help interpret the spatial patterns observed in the condition of biological resources in the estuaries of the Virginian Province. Previous studies have suggested that the incidence of pathological disorders in fish is generally higher in areas with high concentrations of contaminants in the sediment. The species composition of benthic communities also may be affected by high sediment contaminant concentrations; thus, sediment contaminant concentrations can be used as a diagnostic variable for biological response indicators.

A categorical approach can be used to interpret sediment contaminant concentration data as both a diagnostic exposure indicator and an indicator of estuarine condition. Categorical data analysis is appropriate to identify relationships between sediment contamination and biological responses because such responses are unlikely to be linear. Categorical analysis is also appropriate to the needs of environmental managers, whose primary concerns relate to regulatory or target concentrations.

The typical approach for categorical data analysis involves establishing threshold values for the concentration of each contaminant of interest; however, determining appropriate threshold values depends upon the specific question of interest. There are at least three questions of interest to environmental managers:

- Which contaminants are present due to human activities?
- Where are estuarine sediments most contaminated?
- Are sediment contaminants harmful to indigenous biological resources?

The threshold value for addressing each of these questions differs. Although EPA is in the process of developing sediment criteria for estuaries, generally accepted approaches for identifying appropriate threshold values to address these quesitons have not been established.

EMAP is in the process of developing analytical approaches for evaluating sediment contaminant data to address each of the questions. Three potential approaches are presented below, followed in Section 4.2.3.4 by a comparison of the results of applying these methods to evaluate the extent of contamination in estuaries of the Virginian Province. Presenting these methods now does not preclude using other methodologies in the program eventually; evaluation of analytical methods is a continuing activity.

4.2.3.1 Identifying Anthropogenic Enrichment

Sites that have experienced anthropogenic enrichment with organic contaminants can be defined operationally as any sites where there are organic contaminants. Synthetic organics such as pesticides and PCBs only have human origins, and human activities are the main sources of most PAHs. There are natural sources of PAHs, such as the oil seeps in the Southern California Bight, but no such natural sources contribute significant amounts of organic contaminants to the Virginian Province.

Metals in the sediment may be derived from anthropogenic sources or from the natural geochemical processes of weathering and erosion of rocks, since metals occur naturally in the

earth's crust. The difficulty arises in identifying which portion of the total metal content of the sediment is due to natural processes and which is due to human activities. Several methods are used to determine whether measured metal concentrations in estuarine sediments represent natural or anthropogenically-enriched conditions, including analysis of specific grain size fractions (Voutsinou-Taliodouri and Satsmadjis 1982; Ackermann et al. 1983; Schneider and Weiler 1984), normalization by grain-size (NOAA 1990), normalization by organic carbon, and normalization to a reference element (Klinkhammer and Bender 1981; Windom et al. 1989; Trefry et al. 1985; Kouadio and Trefy 1987). Most of these methods were not suitable for EMAP-E. Use of a specific grain size fraction is not practical because EMAP's probability-based sampling design requires using the entire population of samples to generate regional estimates. Similarly, even normalization by grain size requires elimination of sandy samples (NOAA 1990), and one-third of the samples from 1990 were from sandy sediments. Using organic carbon concentrations to adjust sediment metal concentrations was rejected because carbon concentrations are influenced strongly by human activities.

Normalizing concentrations of metals to conservative crustal materials, such as aluminum, has been used to analyze sediment data from selected coastal regions (Bruland et al. 1974; Schropp et al. 1990). The method involves establishing a relationship between the concentration of a particular metal and aluminum. Aluminum-normalized metal concentrations that are significantly greater than the concentrations established by the relationship are interpreted to represent anthropogenically-enriched concentrations. Normalizing to aluminum relies on several characteristics of the metal. Aluminum is highly refractory and does not degrade or change forms in the environment. Metal to aluminum ratios are relatively constant in estuarine sediments without human sources, and aluminum concentrations are not influenced significantly by human activities. Although aluminum normalization techniques have been used to identify enriched sediments over large geographic areas (Windom et al. 1989; Schropp et al. 1990), they have not been used to establish relationships over an area as large as the Virginian Province or with the range of metal concentrations found there.

Simple linear regressions based on log-transformed data were used to determine whether there was a baseline relationship between aluminum and the metals of concern (i.e., chromium, copper, lead, nickel, and zinc) in sediments that are not anthropogenically enriched (Fig. 4-13). Identifying unenriched sediments in the estuaries of the Virginian Province is challenging because of long-standing and widespread contamination. As a first step, sites with significant sediment toxicity were considered enriched and removed from regression calculations. The degree to which this criteria removed all sites already influenced by anthropogenic sources of metals needs to be evaluated. Additional criteria

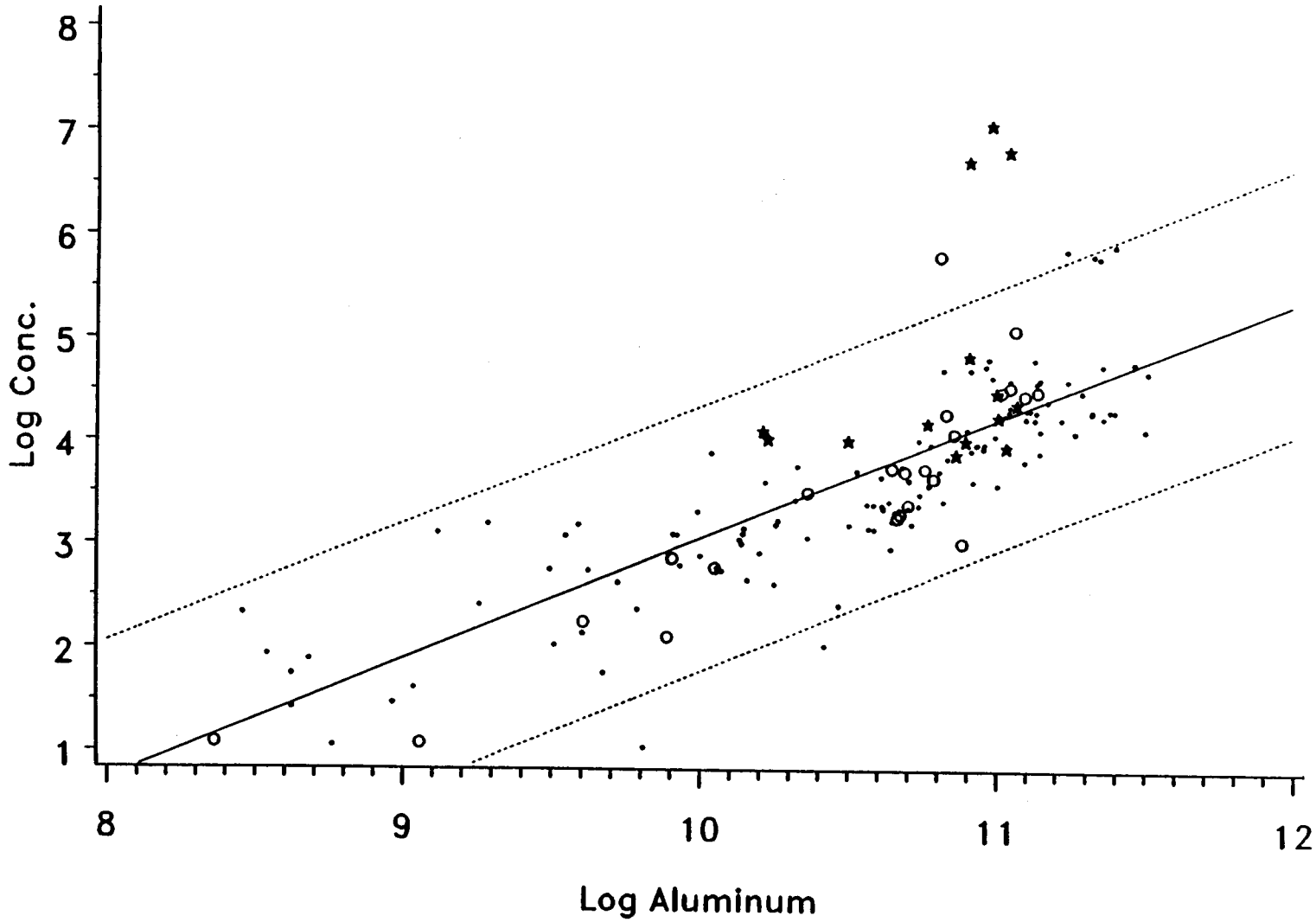

Figure 4-13. Relationship (and 95% confidence intervals) between chromium and aluminum from 1990 EMAP data. Starts indicate site at which the sediments were toxic in the amphipod bioassay.

most likely are needed because there may be other anthropogenically-enriched sites that are not toxic. The slopes of regressions calculated with such sites may be steeper than if calculated only with unenriched sites; consequently, the estuarine area with anthropogenically-enriched metal concentrations may be underestimated. Metal to aluminum relationships and 95% confidence intervals were established for chromium, copper, iron, mercury, manganese, nickel, lead, and zinc. A relationship was not attempted for silver because values above the detection limit were measured at only three sites. Correlation coefficients, slopes, and intercepts for each of the relationships are given in Table 4-13.

Table 4-13. Relationship between sediment metal concentration and aluminum. Correlation coefficient (r^2), slope (m), and intercept (b) values given for the relationship: log (Me^+) =m (log (Al)) + b. Alpha for all relationships except mercury was $\leq$ 0.0001. Alpha for mercury was < 0.002.

Metal	Correlation Coefficient	Slope	Intercept
Cr	0.71	1.042	-7.389
Cu	0.57	1.413	-12.376
Fe	0.75	0.872	0.697
Hg	0.11	0.681	-9.440
Mn	0.40	0.707	-1.332
Ni	0.66	1.095	-8.818
Pb	0.46	0.861	-5.976
Zn	0.63	1.046	-6.528
Cd	0.14	0.613	-7.424

The relationships given in Table 4-13 were compared to those published by Schropp et al. (1990) for uncontaminated sediments from a large area in Florida. Comparisons could be made for chromium, copper, nickel, lead, and zinc. The metal to aluminum relationships were most similar for lead, a heavy metal with a predominantly atmospheric input to estuarine environments. For other metals, the EMAP relationships had greater slopes and smaller intercept values, but the statistical significance of these differences could not be determined without obtaining the Schropp et al. data. This finding suggests that some of the sites used in the EMAP regressions were contaminated by anthropogenic sources of metals. EMAP will

continue to develop aluminum relationships and means for excluding potentially contaminated sites from subpopulations of data used to determine metal-aluminum relationships and for the Virginian Province.

Aluminum normalization relationships also were attempted for mercury and cadmium. These relationships were very weak, though statistically significant, similar to the observation of Schropp et al. (1990). For the EMAP data, aluminum explained only about 10% of the variation in the mercury and cadmium concentrations (Table 4-13). Since the relationship was so poor, any presence of these elements in the sediment was assumed to result from anthropogenic sources. This assumption may be valid for mercury relative to the measurement detection limits of the laboratory procedures used, but it most likely overestimates the anthropogenic contribution of cadmium.

4.2.3.2 Identifying "Above Background" Concentrations

Some contaminants are present throughout the province at relatively low levels. Although the concentrations of these contaminants are high enough to suggest anthropogenic enrichment, the source of enrichment may be atmospheric or diffuse rather than discrete. One interest of managers is to identify "hot spots" resulting from local sources or habitat conditions that concentrate selected chemicals.

Contaminant concentration data for most analyses are log-normally distributed; a high percentage of sample sites has low values, and a lesser percentage has very high values. Selecting an inflexion point to define background in this distribution is a difficult and somewhat arbitrary decision. The most widely used approach for selecting this inflexion point is that of NOAA's NS&T program (NOAA 1990), which uses one standard deviation above the geometric mean.

Applying this approach to EMAP data is problematic; the same data set can not be used to define both the "background," or threshold, concentrations for contaminants and to estimate estuarine condition. An independent data set is required to define threshold concentrations so that the choice of thresholds from the distribution will not directly determine the estimate of estuarine condition. For example, if the EMAP Virginian Province data set was used to define a threshold as the concentration of the 90th percentile of the distribution, then approximately 10% of the sites and 10% of the estuarine area would have high contaminant concentrations.

To resolve this difficulty, sediment data from NOAA's National Status and Trends (NS&T) program were used to define the threshold values. Following the lead of previous analyses

of these data (NOAA 1990), the mean and standard deviation were determined for each contaminant after first transforming the data using a logarithmic function:

$$\text{transformed value} = \ln(\text{concentration} + 1).$$

Logarithmic data transformations made the characteristically log-normal distribution of the contaminant data more gaussian. The "background" value was defined as the mean plus one standard deviation, and the antilog of this sum was used as the threshold value. EMAP sampling sites with contaminant concentrations exceeding these values were classified as having high contaminant concentrations.

4.2.3.3 Determining Biologically Significant Concentrations

Environmental resource managers are concerned with the effects of contaminant concentrations on the condition of indigenous biological resources, regardless of whether the source of the contaminant is natural or anthropogenic. There are no generally accepted sediment quality criteria, despite an extensive literature linking ecological changes from the cellular level to the community level to sediment contaminant concentrations. Various approaches have been used to determine threshold contaminant concentrations for biological effects (Hinga 1992, Long and Morgan 1990; Ginn and Pastorok 1992). These approaches have included relating contaminant concentrations to sediment toxicity test results, benthic infaunal community responses, or both (apparent effects threshold and sediment triad approaches), or have used water quality criteria to determine the toxicity of theoretical pore water contaminant concentrations at equilibrium with sediment contaminant concentrations (equilibrium partitioning approach, USEPA 1990).

Sediment toxicity, benthic infaunal community abundance, and biomass data from the Demonstration Project have not been used to define threshold values for sediment contaminants because the covariance between multiple contaminants was high and the potential effects of individual contaminants could not be separated from the effects of multiple contaminants. Threshold values have been determined from other studies using the various approaches mentioned above. These have merit and will be considered further for application to the EMAP data; however, most examples of threshold values were developed in restricted geographic regions without the range of sediment types, contaminants, or contaminant concentrations represented in the data set for the Virginian Province.

The databases used by Long and Morgan (1990) to compile derived effects values are exceptions. In this investigation, all studies that have demonstrated some type of biological effect were ranked by the concentrations of individual contaminants. Two values were identified for each contaminant: an effects range-low (ER-L) value corresponding to the concentration of the 10th percentile, and an effects range-median value corresponding to the

median (ER-M) concentration, or mid-point, of the ranking. The Long and Morgan ER-L values were interpreted to correspond to threshold contaminant concentrations above which biological effects begin to appear. Except for sensitive species and life stages, these biological effects are likely to be chronic, sublethal effects; therefore, evaluations of sediment contaminant concentrations using the ER-L values provide early warning indicators of pollutant exposure before contaminant concentrations cause widespread acute toxicity. The ER-L values were used in the evaluation of environmental condition of the Virginian Province (Section 2) to estimate the area with contaminants at concentrations having the potential to cause sublethal biological effects; however, as Long and morgan (1990) have noted, acute effects are also possible at ER-L concentrations, particularly to sensitive species and life-stages.

ER-M values were interpreted to correspond to contaminant concentrations above which biological effects are not only possible, but probable. Acute biological effects are more likely to occur at concentrations exceeding ER-M values than at concentrations between ER-L and ER-M values. These interpretations of the Long and Morgan ER-L and ER-M values are similar to those made by Klapow and Lewis (1979), who used a similar analytical approach to assess contaminant concentrations in marine waters. The potential limitation of these interpretations is that the actual effect on biota will depend on the specific taxa present at a site and habitat characteristics of the site that might affect bioavailability of the contaminants.

ER-L and ER-M values are not available for all of the contaminants measured by EMAP (Table 4-14). Values were available for all of the measured metals, 6 of the 15 pesticides, 12 of 23 PAHs, and total PCBs. Contaminants for which there were no ER-L and ER-M values were excluded from the analyses presented in Section 6. This approach may underestimate the number of sites with concentrations of contaminants high enough to cause biological effects. It may also underestimate the prevalence of high organic contamination relative to metals because none of the contaminants for which threshold values are unavailable were metals measured by EMAP. The magnitude of these underestimates appears to be slight, however, since several contaminants in a class for which at least one ER-L value was available generally exceeded that value at a site.

4.2.3.4 Comparison of Analytical Approaches

The threshold values for each contaminant derived by several different methods are compared in Table 4-14. Threshold values are not presented for the anthropogenic enrichment methodologies, since this approach does not employ a single value. For organic contaminants, however, the reader is referred to the median detection limit in Table 4-4.

Table 4-14. Comparison of threshold contaminant levels based upon the "above background", ER-L, and ER-M values. Above background values were defined from NOAA National Status and Trends Sediment Data Base. ER-L and ER-M values taken from Long and Morgan (1990). Values not available for contaminants marked with an asterisk (*).

	"Above Background" Values	ER-L Values	ER-M Values
Metals	**(ppm)**	**(ppm)**	**(ppm)**
Cadmium	0.7	5	9
Chromium	140	80	145
Copper	45	70	390
Lead	46	35	110
Mercury	0.4	0.15	1.3
Nickel	40	30	50
Zinc	150	120	270
Pesticides	**(ppb)**	**(ppb)**	**(ppb)**
Aldrin	0.4	*	*
Alpha-Chlordane	1.0	0.5	6.0
Trans-Nonachlor	1.0	*	*
Dieldrin	1.1	0.02	8.0
DDT - Total	14	3	350
o,p'-DDD	1.4	*	*
p,p'-DDD	4.7	2	20
o,p'-DDE	1.5	*	*
p,p'-DDE	6.1	2	15
o,p'-DDT	0.4	*	*
p,p'-DDT	1.9	1	7
Heptachlor	0.2	*	*
Heptachlor epoxide	0.3	*	*
Hexachlorobenzene	1.1	*	*
Lindane	0.5	*	*
Mirex	0.4	*	*
Polychlorinated Biphenyls	28	50	400

Table 4-14. Continued

	"Above Background" Values	ER-L Values	ER-M Values
Polycyclic Aromatic Hydrocarbons	**(ppb)**	**(ppb)**	**(ppb)**
Acenaphthene	4.5	150	650
Acenaphthylene	10	*	*
Anthracene	26	85	960
Benzanthacene	94	230	1600
Benzo(a)pyrene	100	400	2500
Benzo(e)pyrene	98	*	*
Biphenl	5.3	*	*
Chrysene	133	400	2800
Dibenz(a,b,)anthracene	12	60	260
2,6,-Dimethylnaphthalene	10	*	*
Fluoranthene	229	600	3600
Fluorene	11	35	640
Ideno(1,2,3-c,d)pyrene	42	*	*
1-Methylnaphthalene	10	*	*
2-Methylnaphthalene	19	65	670
1-Methylphenanthrene	22	*	*
Naphthalene	29	340	2100
Perylene	97	*	*
Phenanthrene	113	225	1380
Pyrene	222	350	2200
Benzo(k)fluoranthene	59	*	*
Benzo(g,h,i)perylene	47	*	*
2,3,5-Trimethylnaphthalene	4	*	*

Estimates of the estuarine area with contaminant concentrations of concern were generated using each of the approaches discussed above. Based upon data from the Demonstration Project, almost all of the estuarine area in the Virginian Province showed some evidence of chemical contamination due to anthropogenic activities (Table 4-15). Contamination from organics was most widespread; PAHs were found in nearly all areas (98%). Contaminant concentrations exceeding "background" values were found in 40% of the estuarine area in the Virginian Province. If estuarine areas with high contaminant concentrations were distributed equally around the country, then the expected value, based upon the distribution of the NOAA NS&T data, would have been approximately 17%. The observation that the actual number was over two times larger suggests that the estuaries in the Virginian Province generally have higher contaminant concentrations than estuaries in other parts of the country.

As the question changes from presence or presence above background levels to biological significance, the proportion of estuarine area in the Virginian Province with contaminant concentrations of concern decreases, and metals become of greater concern than organics (Table 4-15). The area with contaminant concentrations having the potential to cause sublethal biological effects (ER-L values) was 39%. In this area, biologically relevant concentrations were due mainly to elevated concentrations of metals (36%). Biologically significant concentrations of organics were more restricted spatially. This finding and comparisons to "background" values suggest that, although contamination from organic compounds may be more pervasive, contamination from metals may have proportionately greater potential for biological effects. This finding also is reflected in analyses using ER-M values as assessment thresholds. The proportion of estuarine area with contaminant concentrations with the potential to cause more acute biological effects (ER-M values) was 7%; most of the elevated concentrations were due to metals.

The extent of contaminant conditions for each of the three classes of estuarine resources also was estimated using the assessment methods outlined above (Figure 4-14). Estimates of contaminated areas varied greatly between the methods. Regardless of method, however, a greater proportion of the estuarine area within large tidal rivers and small estuarine systems had contaminant concentrations of concern when compared to the estuarine area in large estuarine systems.

Following the pattern observed for the province as a whole, nearly all of the estuarine area in all three classes of estuarine resources showed some evidence of chemical contamination due to anthropogenic activities. Sediment contaminant concentrations above background levels were found in 73% of the area in large tidal rivers and 61% of the area in small estuarine systems. Contaminant concentrations above background were less prevalent in large estuaries. A similar pattern was observed using ER-L values as assessment thresholds. A slightly different pattern was apparent using ER-M values. Sediments with contaminant concentrations that are potentially acutely toxic to biota were found in 24% of the estuarine area in small estuarine systems, but were present in less than 5% of the area in large tidal rivers or large estuarine systems (Figure 4-14). These findings suggest that significant sources of contaminants may be present in both large tidal rivers and small estuarine systems. The short residence times of water in large tidal rivers generally minimizes accumulation of extremely high contaminant concentrations in these systems. The longer residence times in small estuarine systems may increase the opportunity for contaminants to become trapped in the sediments, in some cases reaching concentrations sufficient to initiate significant biological effects. The proportionately higher occurrence of sediments with contaminant concentrations exceeding ER-M values is consistent with the finding that sediments that were toxic to laboratory test organisms

Table 4-15. Estimated percent of the area in the Virginan Province exceeding each of four threshold sediment contaminant concentrations.

	Enriched	"Above Background" Values	Exceeding ER-L Values	Exceeding ER-M Values
Contaminants				
All	97.7	40.3	39.1	8.0
Metals				
All	96.3	30.3	36.5	6.9
Cadmium	96.3	13.3	3.1	0
Chromium	4.2	3.1	15.2	3.2
Copper	4.7	14.1	10.2	0.2
Lead	3.3	16.9	27.2	5.2
Mercury	65.3	4.3	21.7	0.7
Nickel	3.5	8.9	15.7	4.3
Zinc	1.4	19.7	24.9	5
Pesticides				
All	22.3	10.1	12.2	1.3
Aldrin	0.7	0	*	*
Alpha-Chlordane	13.0	4.1	5.9	0
Trans-Nonachlor	0.1	0.1	*	*
Dieldrin	5.7	0.6	5.7	0
DDT - Total	5.7	0.5	5.7	0
o,p'-DDD	1.0	0.1	*	*
p,p'-DDD	11.8	0.9	2.1	<0.1
o,p'-DDE	1.7	0.4	*	*
p,p'-DDE	14.4	0.9	4.5	0.4
o,p'-DDT	0.1	0.1	*	*
p,p'-DDT	9.9	3.2	6.2	0.9
Heptachlor	2.0	1.0	*	*
Heptachlor epoxide	0.5	0.1	*	*
Hexachlorobenzene	5.7	0.1	*	*
Lindane	0.3	0.3	*	*
Mirex	5.8	5.8	*	*
Polychlorinated Biphenyls	39.8	6.2	1.0	<0.1

Table 4-15. Continued

	Enriched	"Above Background" Values	Exceeding ER-L Values	Exceeding ER-M Values
Polycyclic Aromatic Hydrocarbons				
All	98.1	27.4	3.7	0.2
Acenaphthene	30.8	15.4	0.2	0.2
Acenaphthylene	19.7	1.1	*	*
Anthracene	33.9	2.8	2.1	0.2
Benzanthacene	85.4	4.1	3.2	0.2
Benzo(a)pyrene	37.7	2.5	0.9	0.2
Benzo(e)pyrene	51.8	3.0	*	*
Biphenyl	24.3	6.3	*	*
Chrysene	81.9	8.4	2.9	0.2
Dibenz(a,b,)anthracene	6.1	4.6	1.1	0.2
2,6,-	21.1	2.5	*	*
Dimethylnaphthalene	69.3	8.3	3.1	0.2
Fluoranthene	30.5	7.9	2.5	0.2
Fluorene	9.3	6.7	*	*
Ideno(1,2,3-c,d)pyrene	45.9	8.5	*	*
1-Methylnaphthalene	62.8	6.4	3.2	0
2-Methylnaphthalene	41.2	3.7	*	*
1-Methylphenanthrene	96.5	15.6	1.3	0
Naphthalene	55.1	1.9	*	*
Perylene	91.2	8.3	3.3	0.2
Phenanthrene	71.9	3.7	3.5	0.2
Pyrene	53.1	5.7	*	*
Benzo(b,k)fluoranthene	19.5	1.9	*	*
Benzo(g,h,i)perylene	6.5	1.8	*	*
2,3,5-Trimethylnaphthalene				

were more prevalent in small estuaries than in either large tidal rivers or large estuarine systems.

The discussion above illustrates that analytical approaches to assessing sediment contaminant concentrations depend greatly on the primary question of interest. Further, assessment results are strongly influenced by method. EMAP intends to work closely with the potential

users of program results to refine assessment methodologies, aiming to identify methodologies that produce the information most relevant to the needs of users.

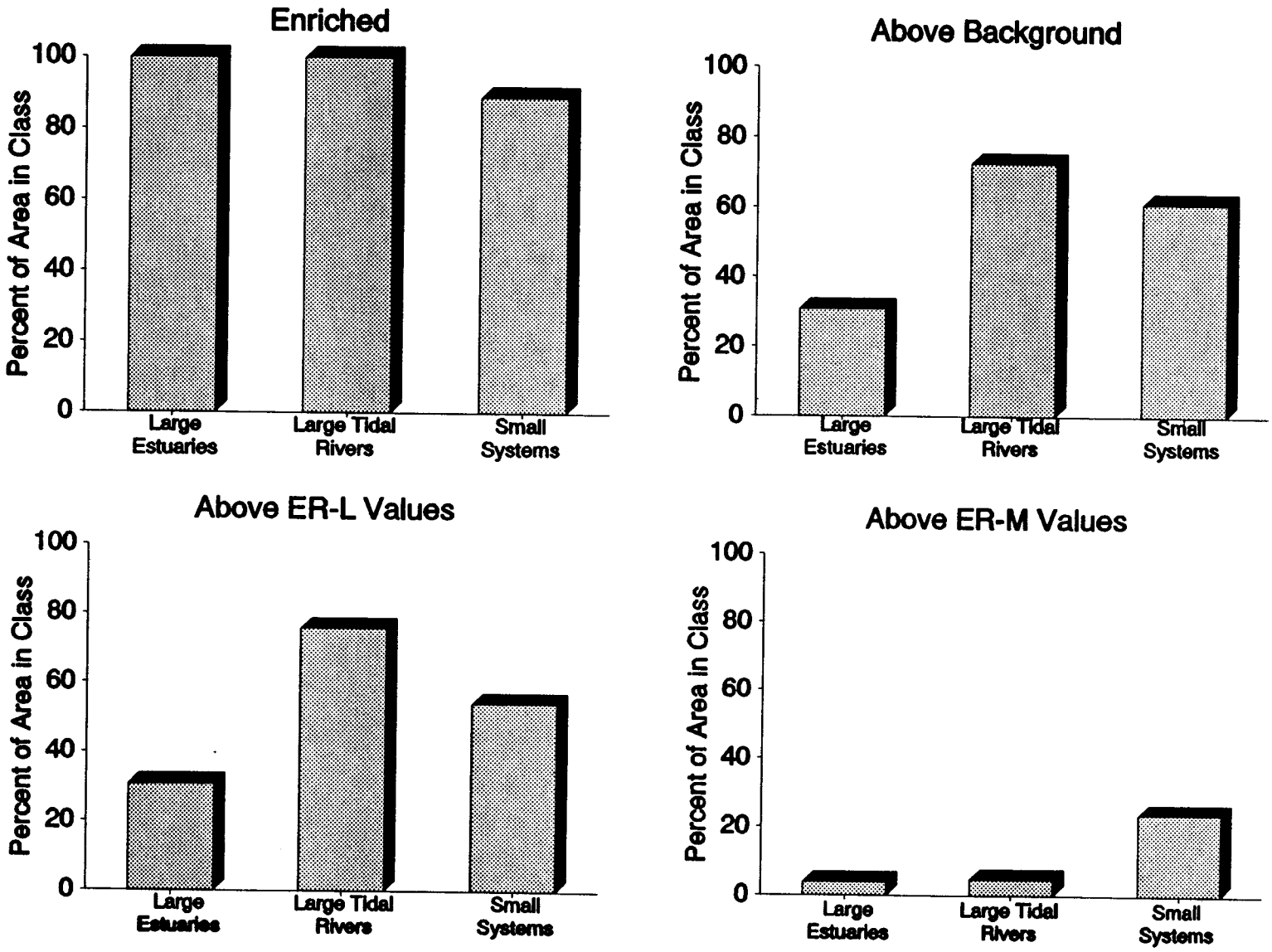

Figure 4-14. Percent of area in each of three resource classes with elevated contaminant concentrations, as defined by four different threshold approaches

4.3 HABITAT INDICATORS

Habitat indicators were not used to distinguish degraded from nondegraded environmental conditions. Instead, habitat indicators were used to describe the physical estuarine environment and to normalize selected response indicators. The three principal habitat indicators were depth, sediment grain size, and salinity.

4.3.1 Depth

Depth was considered to be a stable indicator over the index period and over multiple EMAP sampling cycles. This stability was used to evaluate how well field crews could return to the same sampling site and repeat the same measurement. Separate CDFs were constructed for depth data from sampling intervals 2 and 3 using sampling sites visited during both intervals. The CDFs for these two intervals, which are shown in Figure 4-15, were not statistically different. The degree of similarity between the two CDFs strongly suggests that the methods employed in the 1990 Demonstration Project are adequate to ensure reproducible results.

4.3.2 Salinity

Salinity can vary dramatically at a site within hours as a function of tidal stage, and within weeks as a function of rainfall; consequently, point measurements taken during a single EMAP visit will be inadequate to describe the range of salinity conditions occurring at that site throughout the summer. Similar to dissolved oxygen, however, the salinity profile for the province may remain stable over the summer sampling period. To determine the stability of salinity on a province level, CDF's for intervals 2 and 3 were compared, and no significant differences were found between these two periods (Fig. 4-16). Presumably this results because salinity varies with tide stage at a site, whereas the total salt within the province as a whole remains relatively stable over the summer period. The randomized sampling design, therefore, is an appropriate means for describing conditions in the province, even if it is inadequate for describing the range of conditions at a particular site.

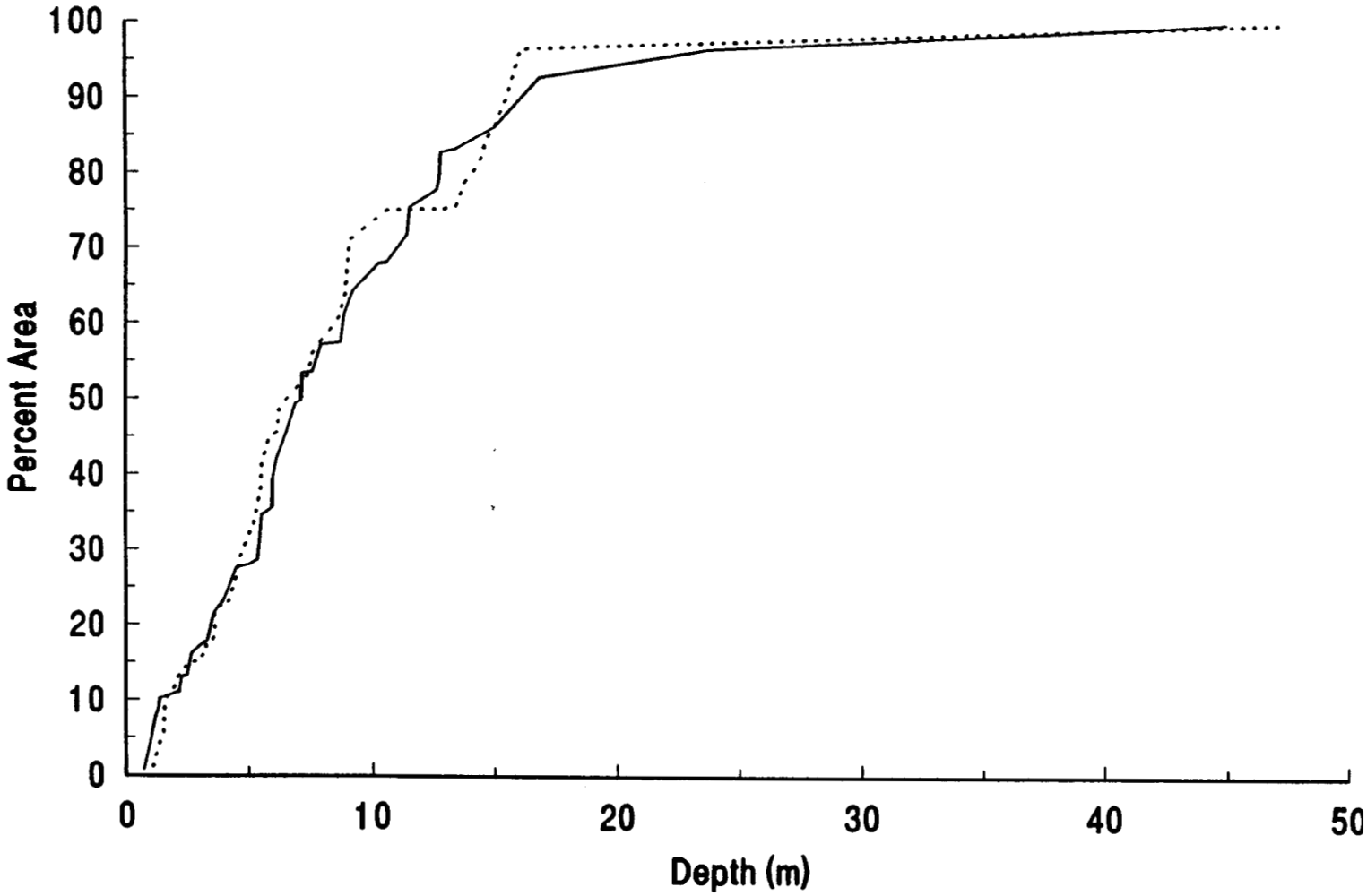

Figure 4-15. Comparison of cumulative distribution function for depth from sampling intervals 2 (solid line) and 3 (broken line)

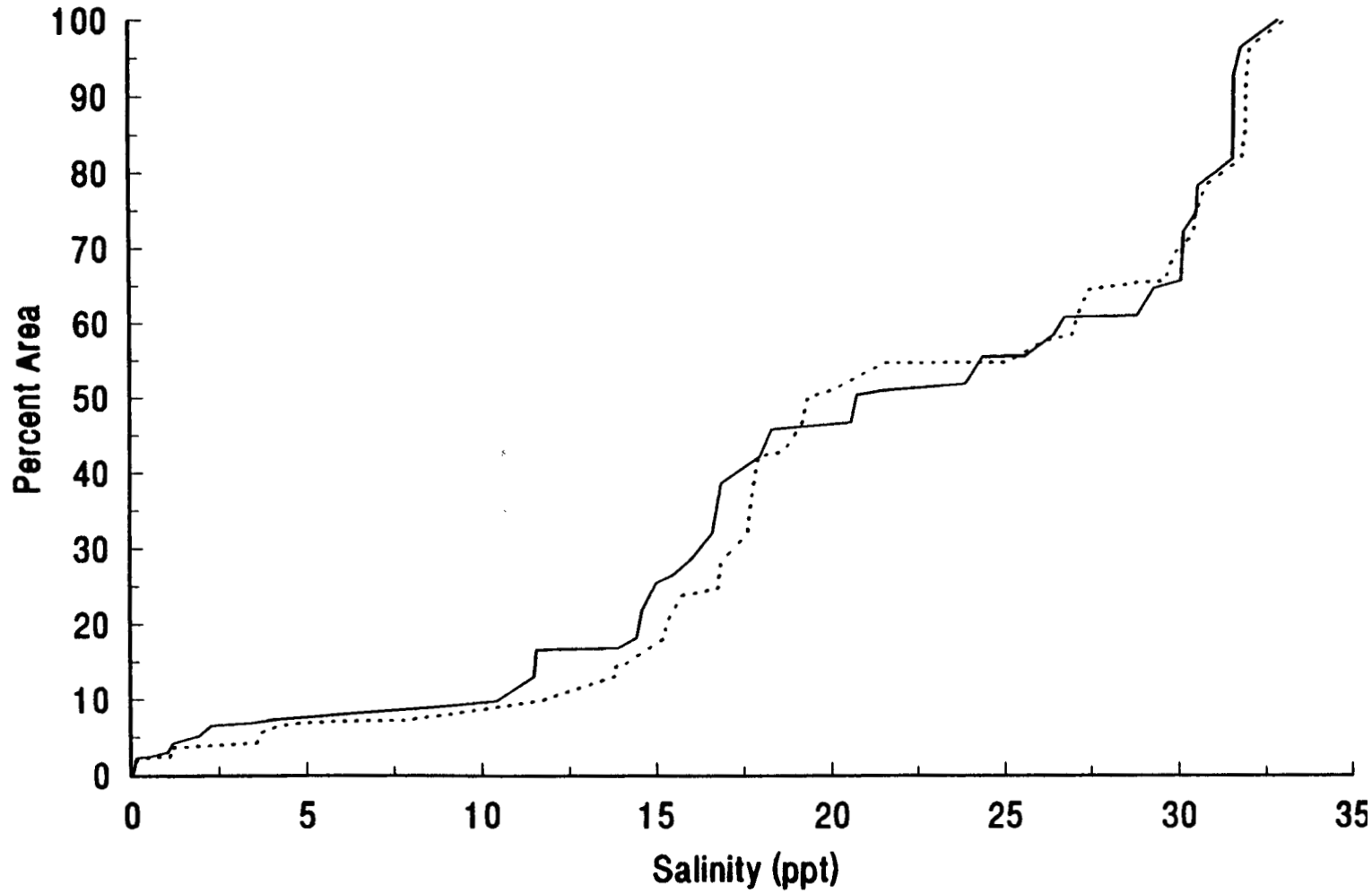

Figure 4-16. Comparison of cumulative distribution function for salinity from sampling intervals 2 (solid line) and 3 (broken line)