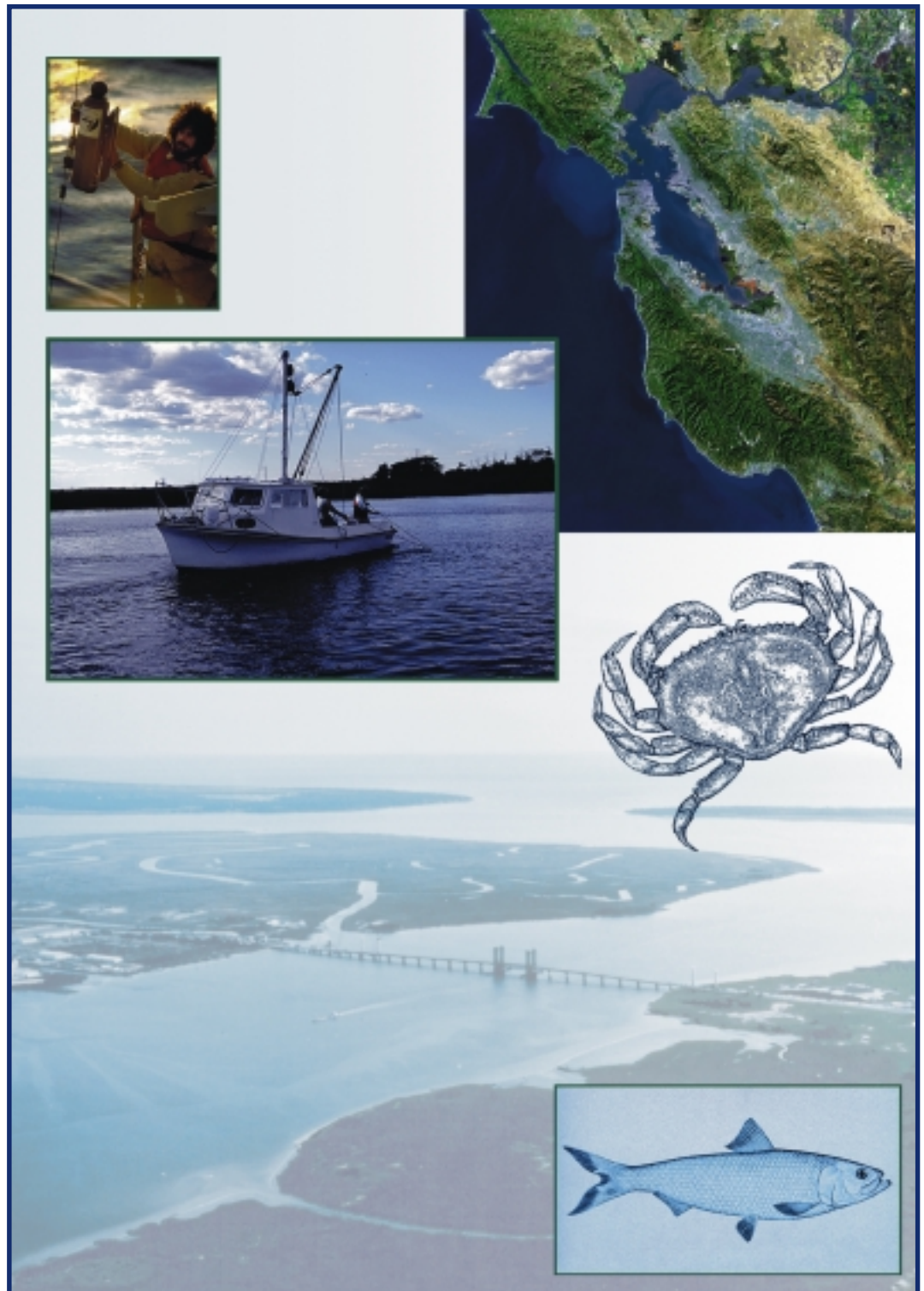**♻EPA**

Estuarine and Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance

Estuarine and Coastal Marine Waters:
Bioassessment and Biocriteria Technical Guidance

George R. Gibson, Jr., Project Leader (4304)
USEPA
Office of Water
Office of Science & Technology
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Principal Authors:

Michael L. Bowman
Principal Scientist
Tetra Tech, Inc.
10045 Red Run Boulevard, Suite 110
Owings Mills, MD 21117

Jeroen Gerritsen
Principal Scientist
Tetra Tech, Inc.
10045 Red Run Boulevard, Suite 110
Owings Mills, MD 21117

George R. Gibson, Jr. (4304)
USEPA
Office of Water
Office of Science & Technology
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Blaine D. Snyder
Senior Scientist
Tetra Tech, Inc.
10045 Red Run Boulevard, Suite 110
Owings Mills, MD 21117

This document is dedicated to the memory of Dr. Douglas Farrell of the Florida Department of Environmental Protection and Dr. Donald Lear, U.S. Environmental Protection Agency (retired). It is fitting that this effort to which they volunteered so much of their invaluable experience and expertise be so dedicated. The benthic community index which Doug developed is also cited here as the "Farrell Index" in further recognition of his unselfish contribution to the protection and management of our coastal resources. Much of the methodology described in the coastal survey portion of this guide was developed from Don Lear's pioneering efforts.

The contributors to this manual sincerely hope that the good common sense, attention to scientific veracity, and practical application of the information to protect our marine resources - so ably personified by Don and Doug - is adequately reflected in these pages.

Disclaimer

This manual provides technical guidance to States, Indian tribes and other authorized jurisdictions to establish water quality criteria and standards under the Clean Water Act (CWA), to protect aquatic life from the effects of pollution. Under the CWA, States and Indian tribes are to establish water quality criteria to protect designated uses. State and Indian tribal decision makers retain the discretion to adopt approaches on a case-by-case basis that differ from this guidance when appropriate and scientifically defensible. While this manual constitutes USEPA's scientific recommendations regarding biological criteria to help protect resource quality and aquatic life, it does not substitute for the CWA or USEPA's regulations; nor is it a regulation itself. Thus, it cannot impose legally binding requirements on USEPA, States, Indian tribes or the regulated community, and might not apply to a particular situation or circumstance. USEPA may change this guidance in the future.

This document has been approved for publication by the Office of Science and Technology, Office of Water, U.S. Environmental Protection Agency. Mention of trade names, products, or services does not convey and should not be interpreted as conveying, official USEPA approval, endorsement or recommendation.

The suggested citation for this document is:

Gibson, G.R., M.L. Bowman, J. Gerritsen, and B.D. Snyder. 2000. Estuarine and Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance. EPA 822-B-00-024. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Acknowledgments

Estuarine and Coastal Marine Biocriteria Workgroup Attendees
1992 - 1997
(Long Term Participation)

Suzanne Bolton
National Marine Fisheries Service
(ST2)
1315 East/West Hwy.
Silver Spring, MD 20910
301-713-2363

Michael L. Bowman
Tetra Tech, Inc.
10045 Red Run Blvd.
Suite 110
Owings Mills, MD 21117

Dan Campbell
c/o USEPA ERL/ORD
University of Rhode Island
27 Tarzwell Drive
Narragansett, RI 02882

Dan Dauer
Department of Biology
Old Dominion University
Norfolk, VA 23529
804-683-3595

Robert Diaz
VA Institute of Marine Science
Glouchester Point, VA 23062
804-642-7364

Cindy Driscoll
MD Department of Natural Resources
Oxford Laboratory
904 S. Morris St.
Oxford, MD 21654
410-226-0078

Charles Eaton
Bio-Marine Enterprises
2717 3rd Avenue North
Seattle, WA 98109
206-282-4945

Larry Eaton
NC Division of Water Quality
4401 Reedy Creek Road
Environmental Sciences Bldg.
Raleigh, NC 27607
919-733-6946

Douglas Farrell*
FL Department of Environmental
Protection
3804 Coconut Dr.
Tampa, FL 33619-8218

Brigette Farren
USEPA, Region III
1650 Arch Street
Philadelphia, PA 19103-2029
215-814-2767

William Muir
USEPA Region III
1650 Arch Street (3ES41)
Philadelphia, PA 19107
215-814-2541

Walt Nelson
USEPA
2111 SE Marine Science Dr.
Newport, OR 97365-5260
541-867-4041

Art Newell
NY Department of Environment
Division of Marine Resources
Building 40, SUNY
Stonybrook, NY 11790-235
516-444-0430

Dave Russell
USEPA Region III
Environmental Science Center
701 Mapes Road
Ft. Meade, MD 20755-5350
410-305-2656

Steve Wolfe
FL Department of Environmental
Protection
2600 Blair Stone Road
Tallahassee, FL 32399

*   Deceased

**Estuarine and Coastal Marine Waters:
Bioassessment and Biocriteria Technical Guidance**

CONTENTS

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

CONTENTS (CONTINUED)

LIST OF FIGURES

Figure

LIST OF FIGURES (CONTINUED)

Figure

LIST OF FIGURES (CONTINUED)

Figure

LIST OF FIGURES (CONTINUED)

Figure

LIST OF FIGURES (CONTINUED)

Figure

LIST OF TABLES

Table

LIST OF TABLES (CONTINUED)

Table

LIST OF TABLES (CONTINUED)

Table

LIST OF TABLES (CONTINUED)

Table

Acronym List

APHA	American Public Health Association
AVS	Acid Volatile Sulfides
BMP	Best Management Practices
CCA	Canonical Correspondence Analysis
CDF	Cumulative Distribution Function
CSREES	Cooperative State Research, Education, & Extension Service
CTD	Conductivity - Temperature - Depth Meter
CV	Coefficient of Variation
CWA	Clean Water Act
DFA	Discriminant Function Analysis
DGPS	Differential Global Positioning System
DMRs	Discharge Monitoring Reports
DNR	Department of Natural Resources
DO	Dissolved Oxygen
EMAP	Environmental Monitoring & Assessment Program
EPA	Environmental Protection Agency
ER-L	Effects Range-Low
ER-M	Effects Range-Median
FEI	Farrell Epifaunal Index
FTE	Full Time Equivalent
GIS	Geographic Information System
GPS	Global Positioning System
IBI	Chesapeake Bay Estuarine Index of Biotic Integrity
ITI	Infaunal Trophic Index

MDS	Multidimensional Scaling
NMDS	Non-metric Multidimensional Scaling
NMFS	National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center
NPDES	National Pollutant Discharge Elimination System
NS&T	National Status & Trends
PAHs	Polyaromatic Hydrocarbons
PCA	Principle Components Analysis
PCBs	Polychlorinated Biphenyls
PCE	Power Cost Efficiency
PCS	Permit Compliance System
POTW	Publically Owned Treatment Works
QA	Quality Assurance
QC	Quality Control
RBP	Rapid Bioassessment Protocol
RPD	Redox Potential Discontinuity
SAV	Submerged Aquatic Vegetation
SEM	Simultaniously Extracted Metals
SOP	Standard Operating Procedure
SPM	Suspended Particulate Matter
SQG	Sediment Quality Guidelines
SQT	Sediment Quality Triad
STORET	STOrage & RETrieval
TDN	Total Dissolved Nitrogen
TDP	Total Dissolved Phosphorus

TMDL	Total Maximum Daily Loads
TOC	Total Organic Carbon
TPC	Total Particulate Carbon
TPN	Total Particulate Nitrogen
TPP	Total Particulate Phosphorus
TSS	Total Suspended Solids
TVS	Total Volatile Sulfides
TWINSPAN	Two-Way INdicator SPecies ANalysis
UPMGA	Unweighted Pair Group Mean Averages
USDA CSREES	United States Department of Agriculture Cooperative State Research Education Extension Service
USGS	United States Geological Survey

Executive Summary

This technical guidance document is based on the concept that bioassessment and biocriteria programs for estuaries and near coastal waters are interrelated and critical components of comprehensive water resource protection and management. Understanding how estuarine ecosystems function and respond to human activity requires a holistic approach to protection and management that integrates biological assessments into the more traditional chemical and physical evaluations. Section 101 of the Clean Water Act requires federal and state agencies to "restore and maintain the chemical, physical, and biological integrity of the nation's waters." Relatively undisturbed aquatic ecosystems have high *biological integrity*, defined as

> *the condition of an aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by an evaluation of multiple attributes of the aquatic biota. Three critical components of biological integrity are that the biota is (1) the product of the evolutionary process for that locality, or site, (2) inclusive of a broad range of biological and ecological characteristics such as taxonomic richness and composition, and trophic structure, and (3) is found*

in the study biogeographic region (USEPA 1996a)[1]

In water resource monitoring and protection, biological criteria are an important addition to the traditional physical and chemical criteria used by EPA. The relative biological integrity, or quality, of the resource can be assessed by comparing the health and diversity of its biological communities to the health and diversity of biological communities in waters with the same physical characteristics but which are relatively unimpacted by human development . There are basically four elements that comprise biocriteria:

1. Reference waters (relatively undisturbed areas that can be compared to study areas) serve as "benchmarks" of water resource quality decision making.

2. The historical record of the biological quality, diversity and productivity.

3. Model projection of the historical and reference condition data (if necessary).

4. The objective assessment of this information by a regional panel of specialists such as state,

---

[1] *Biological criteria: Technical guidance for streams and small rivers.* EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

academic, and federal estuarine ecologists, chemists, fisheries biologists, oceanographers, and resource managers.

The summation of these four factors is the biological criterion for a given estuary or class of coastal water in a geographic region. Examples of the parameters included in a biocriterion are community measures or indexes drawn from dynamic assessments of resident fish, benthic invertebrate, macrophyte, and planktonic assemblages making up the biological community.

Many natural resource agencies throughout the United States have begun the process of developing and implementing bioassessments and criteria programs primarily for rivers and streams. This document is part of the effort to advance the use of these strategies with regard to estuaries and near coastal waters, thereby fostering the development of credible and practical bioassessment programs. This document is intended to provide managers and field biologists with functional methods and approaches for bioassessment and biocriteria development.

In developing biological information, it is imperative that the physical and chemical habitat be carefully measured and documented. Information such as salinity, depth, sediment grain size, and water quality (including pH, temperature, DO, nutrients, and toxicants) is essential to proper classification of the waters for comparison and to the potential subsequent investigation of possible

causes of degradation so that responsible management can be initiated.

This guidance provides detailed descriptions of the appropriate habitat measurements to make the subsequent physical classification to be achieved. The document then describes four levels of investigative intensity or sampling tiers. These tiers are suggested as one possible approach to organizing the data gathering efforts and investigation needed to be able to establish biocriteria in a scientifically defensible manner. Other approaches using variations of these tiers may be appropriate depending on program objectives.

▸ Tier 0 is a preliminary review of existing literature and data available for the estuary or coastal water of concern. It provides candidate reference sites for the development of a reference condition;

▸ Tier I is a one-time site visit with preliminary data gathering to refine the information in Tier 0 and establish candidate biocriteria;

▸ Tier II repeats and builds on measurements initiated in Tier I and establishes the reference condition data which is combined with the historical record, possible models or other extrapolations, and a consensus of regional expert opinion to establish and employ the biocriteria for management decision making;

▸ Tier III is the diagnostic investigation requiring the most

sampling events and most extensive parameters to help establish management efforts for those waters which do not meet the biocriteria.

Biocriteria development is **not** a *one size fits all* proposition. Biocriteria can be developed on biogeographical province basis or on a smaller local basis to account for the geographic, climatologic, and biologic variation in the country. Reference conditions and biocriteria must be specific to each part of the country in order to be responsive and useful for decision making. It is important to remember that such circumstances vary and that this document cannot address every situation or experience. It is oriented toward practical decision making rather than research. Its primary audience is intended to be state and tribal resource managers. It is also intended to provide managers and biologists with functional methods and approaches to facilitate the implementation of viable bioassessment and biocriteria programs that meet their individual needs and resources.

Biocriteria can be used to help support and protect designated uses of water resources; expand and improve water quality standards; detect problems other water quality measurements may miss or underestimate; help water resource managers set priorities for management planning and, assess the relative success or failure of management projects.

Biocriteria do not supersede or replace physical or chemical criteria for water resource decision making and management. In fact biocriteria augment these established measures so

USEPA and the States and Tribes are better informed about the quality of our nations extensive and coastal water resources. The bioassessment/ biocriteria process is a particularly cost effective screening tool to evaluate over all water quality and determine water resource status and trends. The following table shows the progression of the biocriteria process.

Sequential progression of the biocriteria process. Adapted from
Paulsen et al. 1991.

Step 1	**Preliminary Classification to Determine Reference Conditions and Regional Ecological Expectations** • Resource classification • Determination of best representative sites (reference sites representative of class categories)
Step 2	**Survey of Reference Sites and Selected Impaired Sites** • Collection of data on biota and physical habitat • Compilation of raw data (taxonomic lists, abundance levels, and other direct measures and observations)
Step 3	**Final Classification** • Test preliminary classification • Revise if necessary
Step 4	**Metric Evaluation and Index Development** • Data analysis (data summaries) • Testing and validation of metrics by resource class • Evaluation of metrics for effectiveness in detecting impairment • Selection of biological endpoints • Aggregation of metrics into index. • Test the index for validity on another data set.
Step 5	**Biocriteria Development** • Adjustment by physical and chemical covariates • Adjustment by designated aquatic life use
Step 6	**Implementation of Monitoring and Assessment Program** • Determination of temporal variability of reference sites • Identification of problems
Step 7	**Protective or Remedial Management Action** • Initiate programs to preserve exceptional waters • Implement management practices to restore the biota of degraded waters and to identify and address the causes of this degradation
Step 8	**Continual Monitoring and Periodic Review of References and Criteria** • Biological surveys continue to assess efficiency of management efforts • Evaluate potential changes in reference condition and adjust biocriteria as management is accomplished

Chapter 1

Introduction: Bioassessment and Biocriteria

1.1   Rationale

1.1.1   Water Quality Monitoring

The recognition that chemical water quality analyses do not adequately predict or reflect the condition of all aquatic resources has led to the development of measures of biological integrity expressed by biological criteria. Biological surveys, criteria, and assessments complement physical and chemical assessments of water quality by reflecting the cumulative effects of human activities, and natural disturbances on a water body, including the possible causes of these effects. The biological approach is best used for detecting generalized and non-specific impairments to biological integrity, and for assessing the severity of those impairments. Then, chemical and toxicity tests, and more refined habitat assessments, can be used to identify probable causes and their sources, and to suggest corrective measures.

For the purposes of bioassessment and biocriteria development described here, an estuary is a semi-enclosed water body that has a free connection with the open sea and an inflow of freshwater that mixes with the seawater; including fjords, bays, inlets, lagoons, and tidal rivers. Coastal marine waters are those marine waters adjacent to and receiving estuarine discharges and extending seaward over the continental shelf and/or the edge of the U.S. territorial sea.

1.1.2   Advantages of Bioassessment and Biocriteria

Bioassessment is intended to detect biological responses to pollution and perturbation. Routine water quality monitoring for example, detects effects of nutrient enrichment and chronic acidification, but normally is not designed to detect trace levels of toxicants or contaminants, ephemeral pollution events (e.g., acidic episodes, spills, short-lived toxicants and pesticides, short-term sediment loading), or combined or synergistic impacts. Bioassessment, by monitoring organisms that integrate the effects of environmental changes, may in time detect these effects.

Bioassessment, coupled with habitat assessment; i.e., physical and chemical measurements, helps identify probable causes of impairment not detected by physical and chemical water quality analyses alone, such as nonpoint source pollution and contamination, erosion, or poor land use practices. The detection of water resource impairment, accomplished by comparing biological assessment results to the biological criteria, leads to more definitive chemical testing and investigations which should reveal the cause of the degradation. This, in turn, should prompt regulatory and other management action to alleviate the problem. Continued biological monitoring, with the data collected compared to the criteria, will determine the relative success of the management efforts.

1.2 Legal Origins

1.2.1 Clean Water Act

The CWA, Section 101, requires federal and state governments to "restore and maintain the chemical, physical, and biological integrity of the nation's waters." Thus, the Act mandates the restoration and maintenance of biological integrity in the Nation's waters. The combination of performing biological assessments and comparing the results with established biological criteria is an efficient approach for evaluating the biological integrity of aquatic ecosystems. Other pertinent sections of the CWA are Sections 305(b), 301(h), and 403(c). Table 1-1 outlines suggestions for the application of biological monitoring and biocriteria for estuaries through existing state programs and regulations.

1.2.2 305(b) Reporting

States and the USEPA report on the status and progress of water pollution control efforts in §305(b) reports submitted every two years. Inclusion of biological assessment results in these reports will improve the public understanding of the biological health and integrity of water bodies. Many of the better known and widely reported recoveries from pollution have involved the renewal or reappearance of valued species to systems from which they had nearly disappeared, or the recovery of a viable fishery from contaminants. Examples of such recoveries are the restoration of the lower Potomac River and of shellfish beds in Maine. Incorporation of biological integrity in §305(b) reports will ensure the inclusion of a bioassessment endpoint, and will make the reports more accessible and meaningful to many segments of the public.

1.2.3 301(h) and 403(c) Programs

Two other programs within USEPA that specifically rely on biological monitoring data in coastal marine areas are the §301(h) Waiver Program and the §403(c) Ocean Discharger Program. The §301(h) program allows estuarine and marine dischargers who meet specific criteria set forth by USEPA to defer secondary treatment if they can show that their discharge does not produce adverse effects on resident biological communities. As part of the modified NPDES permit received through this waiver program, the dischargers are required to conduct extensive biological monitoring programs designed to detect detrimental effects to those biological communities.

The §403(c) Ocean Discharge Program requires that all dischargers to marine waters provide an assessment of discharge impact on the biological community in the area of the discharge and on the surrounding biological communities. This program requires extensive biological monitoring for some dischargers. Community bioassessment methods are valuable in this program for trend assessment and, in some cases, refinement into more rigorous and definitive assessments.

1.2.4 304(a) Criteria Methodology

This technical guidance was developed under the §304(a) requirement that, "criteria for water quality accurately reflecting the latest scientific knowledge of the kind and extent of all identifiable effects on health and welfare including, but not limited to, plankton, fish, shellfish, wildlife, plant life, shorelines, beaches, aesthetics, and recreation which may be expected from the presence of pollutants in any body of

water . . ." be published and updated as needed.

Under this section, a guidance document must include information on restoration and maintenance of chemical, physical, and biological integrity of navigable and ground waters, waters of the contiguous zone, and the ocean. This also covers information identifying conventional pollutants, such as those classified as biological oxygen demanding,

Table 1-1. Applications of estuarine biological monitoring protocols and biocriteria.

Program	Biological Monitoring and Assessment	Biological Criteria
Section 305(b)/ Reporting	• Improving data for beneficial use assessment. • Improving water quality reporting.	• Identifying waters that are not achieving their aquatic life use support. • Defining an understandable endpoint in terms of "biological health" or "biological integrity" of waterbodies."
National Estuary Program (NEP)	• Assessing status of biological components of estuarine systems. • Develop monitoring objectives and performance criteria. • Establish testable hypothesis and select statistical methods. • Assessing estuarine trophic status and trends, and assessing biological trends. • Select analytical methods & alternative sampling designs. • Evaluate expected monitoring study performance. • Implement monitoring study & data analysis. [Monitoring and sampling needs vary for each estuary]	• Identifying estuaries that are not attaining designated use (including aquatic life use) support. • Defining estuarine biological integrity based on a reference condition. • Identifying impairments due to toxic substances, eutrophication, and habitat modification.
Section 319/Nonpoint Source Program	• Evaluating nonpoint source impacts and sources. • Measuring site-specific ecosystem response to remediation or mitigation activities. • Assessing biological resource trends within watersheds.	• Determining effectiveness of nonpoint source controls.
Watershed Protection Approach	• Assessing biological resource trends within watersheds.	• Setting goals for watershed and regional planning.
TMDLs	• Identifying biological assemblage and habitat impairments that indicate nonattainment of water quality standards. • Priority ranking waterbodies. • Documenting ecological/water quality response as a result of TMDL implementation.	• Identifying water quality-limited waters that require TMDLs. • Establishing endpoints for TMDL development, i.e., measuring success.

Program	Biological Monitoring and Assessment	Biological Criteria
NPDES Permitting	• Measuring improvement or lack of improvement of mitigation efforts. • Developing protocols that demonstrate the relationship of biological metrics to effluent characteristics.	• Performing aquatic life use compliance monitoring. • Helping to verify that NPDES permit limits are resulting in achievement of state water quality standard.
State Monitoring Programs	• Improving water quality reporting. • Documenting improvement or lack of improvement of mitigation efforts including estuary clean-up efforts, TMDL application, NPDES efforts, nonpoint source pollution controls, etc. • Problem identification and trend assessment. • Prioritizing waterbodies.	• Providing a benchmark for measuring effectiveness of controls and performing watershed/regional planning.
Risk Assessment	• Providing data needed to estimate ecological risk to assessment endpoints.	• Providing an assessment or measurement endpoint.
Water Quality Criteria and Standards	• Developing data bases for estuarine phytoplankton, macroinvertebrates, fish, plants, and other assemblages. • Developing indices that assess estuarine biota compared to a reference. • Providing data for aquatic life use classifications.	• Providing benchmark for identifying waterbodies that are not attaining aquatic life use classification. • Developing site-specific standards.
Section 301(h)/ Waiver Program	• Allows marine discharges who meet USEPA criteria to defer secondary treatment if discharge does not produce adverse effects on resident biological communities.	• Providing threshold against which to measure detrimental effects on biological communities.
Section 403(c)/Ocean Discharge Program	• Requires marine dischargers to provide an assessment of discharge impact on biological community in discharge area as well as surrounding communities.	• Providing threshold against which to measure discharger impacts on biological communities.
Section 304(a)/ Criteria Methodology	• Provides information on restoration and maintenance of chemical, physical, and biological integrity of waters. • Identifies conventional pollutants, their concentrations and effects on surrounding communities.	• Providing the benchmark for measuring the effects of pollutants on the biological community.

suspended solids, fecal coliform, and pH. Section 304(a)(8) authorizes USEPA to develop and publish methods for establishing and measuring water quality criteria for toxic pollution, on other bases than a pollutant by pollutant approach. This includes biological monitoring and assessment methods. Specific states have the authority to enforce more stringent regulations as necessary.

1.2.5 Biocriteria

A major purpose of developing biological assessment methods is to establish biological criteria for surface waters. Biological criteria are guidelines or benchmarks adopted by states to evaluate the relative biological integrity of surface waters. The criteria are defined as "narrative expressions or numerical values that describe the biological integrity of aquatic communities inhabiting waters of a given designated aquatic life use" (USEPA 1990). Biological criteria are, in effect, a practical approach to establishing management goals designed to protect or restore biological integrity. Biocriteria can be adopted by a State into their water quality standards, along with chemical, physical and toxicity criteria to better protect aquatic life uses of waterbodies.

Biocriteria can be developed from reasonable expectations for the locality based on: historical data; reference conditions; empirical models; and the consensus judgment of regional experts (Section 1.4.2). The reference condition component of biocriteria requires minimally impaired reference sites against which the study area may be compared. Minimally impaired sites are not necessarily pristine; they must, however, exhibit minimal influence by human activities relative to the overall region of study (USEPA 1996a). In some

instances, "minimally impaired" sites are not available because the entire area has been degraded. Biocriteria are then based on historical data, empirical models if appropriate, and expert judgement to set a condition better than present sites. Restoration of the degraded area must therefore be accomplished before any such reference sites can be established.

Biological criteria typically include the condition of aquatic communities at designated reference sites as an important component. The conditions of aquatic life found at these sites are used to help detect both the causes and levels of risk to biological integrity at other sites of that type in a region. In keeping with the policy of not degrading the resource, the reference conditions—like the criteria they help define—are expected to be upgraded with each improvement to the water resource. It is important that biological criteria not be based on data derived from degraded reference sites. In fact, a concerted effort should be made by States and other jurisdictions to preserve the quality of designated reference sites by setting those areas aside in preserves or parks or by inclusion in use protection programs so that continuity of the biocriteria data base can be maintained. Biocriteria supported by bioassessment surveys serve several purposes in surface water programs, discussed in the following section.

1.3 Uses of Biocriteria

The biocriteria-bioassessment process helps resource managers identify impairment of designated beneficial uses. It expands and improves designated beneficial use classifications and their associated water quality standards. It detects problems other

survey methods may miss or underestimate. It is a process which helps the resource manager set program priorities. It can also be used to evaluate management and regulatory efforts. For example, the information summarized in Table 1-2 indicates that wastewater outfalls are a controlling factor of soft bottom benthic communities and that there is a moderate scientific understanding of the effects of these outfalls specifically in the Southern California Bight (USEPA 1992).

1.3.1 The Use of Bioassessment Data to Establish Biocriteria Appropriate to Designated Beneficial Uses

The hypothetical information presented in Figure 1-1 represents data collected for a given class of similar estuarine or coastal reaches (e.g., similar sediments, depths, and salinities) within the same geographic region. For these areas some high level of resource quality can be conceived which represents a pristine condition, essentially the optimum potential or integrity of those waters. A completely unimpaired (no negative human impacts upon the organisms of the natural system) estuary or coastal marine area is referred to as having biological integrity. The approximation of this ideal quality at the top of a continuum can be expressed by a variety of environmental measures of the biota indicated on the vertical axis of the graph. The determined **ideal** level of biological measurements at the maximum score is shown by the upper horizontal line (equivalent to biological integrity). A second horizontal line somewhat below this is the level set as the reference condition, the **attainable** level of integrity derived from actual measurements from among the highest quality areas in the class. All information on this axis is expected to be objectively derived through the scientific

process and usually is presented in a comprehensive index of many biological characteristics such as an IBI or the EMAP benthic index (Chapter 11).

The horizontal axis represents a progression of socially determined **use designations**; i.e., those predominant uses the State has concluded are appropriate for a particular estuary, region or area within the class. These hypothetical designated uses are arranged on the graph from those usually associated with relatively low water resource quality on the left, to those associated with very high, relatively natural, resource quality on the far right.

The potentially optimal array of **biological criteria** for this class of waters, then, are scores between the reference condition and the level of biological integrity; i.e., between that which is achievable and that which is ideal. The narrower this area, the higher the quality of the waters throughout the class, and the less restoration management is required. The objective, then, is to protect these resources.

On the same horizontal axis, a class of high quality regional uses are further described by a subset of **aquatic life uses**. These are the designated uses for which management goals are also described by desirable characteristics of the aquatic biota to be especially protected, such as "protection of the health and diversity, undiminished, of all indigenous species of fish and invertebrates" for those designated as exceptional natural waters. Resource managers need to apply their first, concerted efforts to those uses because it is usually more cost-effective and resource-conservative to protect existing high quality areas than it is to restore degraded ones.

Introduction - Bioassessment & Biocriteria

Table 1-2. Impacts on the marine environment of the Southern California Bight. Modified from Bernstein et al. 1991.

Sources of Perturbation	Intertidal	Phytoplankton	Zooplankton	Soft Bottom Benthos	Hard Bottom Benthos	Kelp Beds	Wetlands & Estuaries	Commercial Shellfish	Pelagic Fish	Demersal Fish	Fish Eggs & Larvae	Marine Mammals	Marine Birds	Human Health	All
Storms	△			▲	●	☆	◉						○		
El Niños	▪	★	★	△	○	☆	△		★	▲		○	▲		
California Current		★	★						★	★	☆				
Upwelling		★	★			☆			★		★		●		
Blooms/Invasions				△		◉			◉	◉					
Ecol. Interactions	★	★	★	●		★	○	▲	▢		△		▢		
Power Plants								●	●		○				
Wastewater Outfalls		○		☆	●	○		○		▪		○	▢	○	
Dredging				◉		○	▢								
Rivers/Storm Runoff	△					▢	☆	○		○	●	●	▢		
Commercial Fishing						△		★	△	▪		◉	◉	○	
Sport Fishing									▪	▪	▪	○		▪	
Habitat Loss/Mod.	☆					▢	☆	○		○	●	●	▢		
Oil Spills	☆						▲						▢	○	
All						Net effect of each source on all components									

Right margin (vertical): Net effect of each source on all components

KEY		
Potential Importance		**Understanding**
☆ Controlling ▢ Moderate		■ High
△ Major ○ Some		▨ Moderate
		▢ Low

Figure 1-1

Biocriteria for given classifications of estuaries and coastal marine areas. Shaded boxes represent the appropriate biocriterion range for selected classes. Unshaded boxes represent the range of measurement results for test sites in given classes. The vertical arrows above the boxes for the "significantly altered estuaries and coastal marine areas" class indicate the goal of raising the biocriterion for these waters over time in response to restoration efforts.
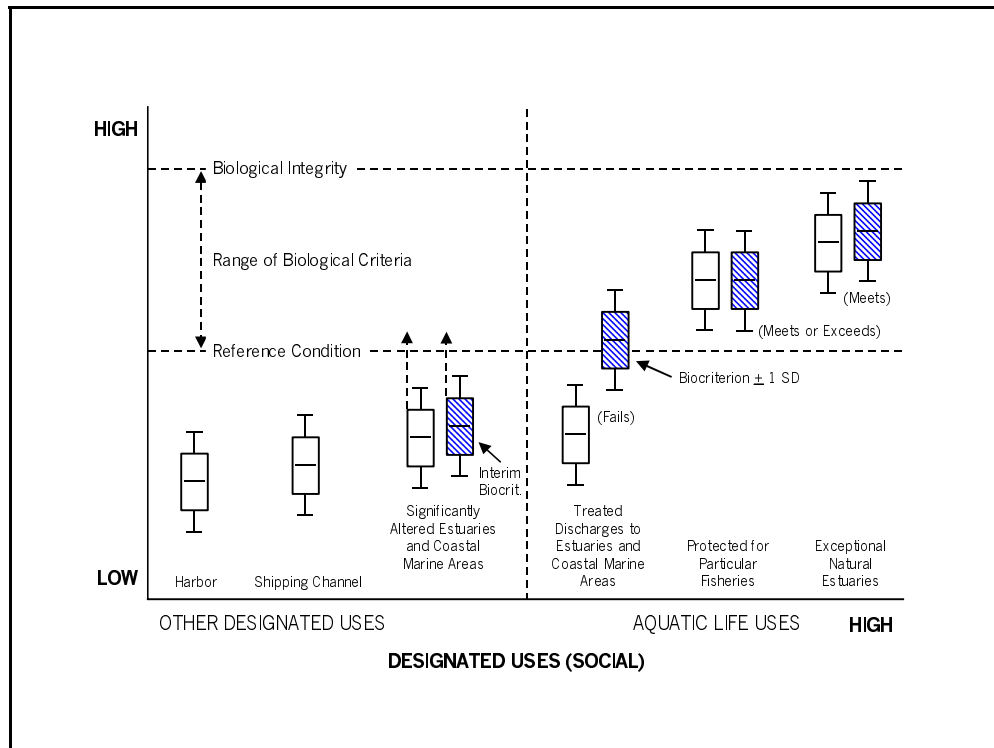
Selected biocriteria with an acceptable range of variation, perhaps one standard deviation, are shown as cross hatched boxes appropriately located for each designated use. Test results for a given area in any use classification ("box and whisker" plots showin g the full range of measurements including variation for that area) can then be compared graphically to the biocriterion for that designated use. Three interpretations of an estuarine or coastal marine area meets its criterion, meets or perhaps even exceeds its criterion, and fails to meet the criterion are illustrated.

A fourth possible result is the marginal condition of significantly altered systems such as urban harbors or shipping channels. The original condition of these areas may very well have been within the optimal range of biotic health and diversity for the region, but intense development has significantly altered them so that as a group they no longer meet the minimum reference condition

for similar areas of the region. An interim biocriterion for these areas may be set with the intention of progressively raising the criterion when sequential restoration efforts are accomplished through a long range management effort.

The "other desi gnated uses" to the left of the bifurcation line may still be surveyed to assist management decision makers; however, they fail to meet the criteria, and there are no designated aquatic life uses which apply.

The designated uses, aquatic life uses and biocriteria are all hypothetical in this illustration, but the interrelationships of societal and scientific elements of decision making should be evident. They are independent processes linked by an environmental ethic and the USEPA policy of antidegradation of water resource quality (the reference condition "bottom line" so to speak). A rational decision can be made which balances that which is ideal with that which is

achievable measured by the objective processes of science.

1.3.2 Expansion and Improvement of Water Quality Standards

When a State adopts biological criteria in their water quality standards to protect aquatic life uses, the criteria become benchmarks for decision making, and may form the basis for requirements in NPDES permits and other regulatory programs.

1.3.3 Detection of Problems Other Methods May Miss or Underestimate

In the process of establishing biocriteria, more data and information is inevitably developed than was previously available. The review of this new information often reveals problems not evident before or provides expanded insight into existing concerns and issues. Armed with this information, a water resources manager is better able to examine issues and make decisions.

1.3.4 Helping the Water Resource Manager Set Priorities

In light of the new information described above, the schedule of activities, allocation of funds, and uses of personnel and equipment may be more appropriately prioritized according to the urgency or magnitude of the problems identified.

With the expanded available biological information augmenting chemical and physical information, managers can apply a triage approach to water resource projects based on the actual condition of the biota affected. This is much like a physician evaluating multiple emergency medical patients. Essentially, areas that are critically impaired, those that are moderately impaired, and those in good condition for which protection rather than remediation is required, can all be identified. Rational decisions can then be made about how to apply limited resources for the best results in accordance with the needs and priorities of the state.

1.3.5 Use of Biosurveys and Biocriteria to Evaluate the Success or Failure of Management Initiatives or Regulations

The manager may design a biosurvey to collect data before and after a permit, regulation or other management effort has been implemented, perhaps augmented by spatially distributed nearfield/farfield sampling as well. With this information and the biocriteria decision making benchmark, it is possible to clearly evaluate the environmental response of the system to the methods applied. This is useful in the NPDES permit review process as a way to help determine the effectiveness of permit controls. Typically, biocriteria are not used directly in NPDES permits as effluent limitations. Biomonitoring above and below a permit site when compared to the established biocriteria will reveal the adequacy of the permit to achieve its intended purpose.

If the biota are unimpaired or recovering, it may be wise to leave the permit, management practice or regulation as is. If the biota are impaired or declining, the review recommendation may be to change the permit, management technique or regulation accordingly. With NPDES permits, the five year review cycle allows sufficient time for extensive biological information to be developed so this determination can be made with reasonable confidence.

1.4 Program Interdependence

It should be readily evident from the applications described above that physical, chemical, and biological surveys and monitoring (repetitive surveys of the same area) and biological criteria are interrelated in the water resource management process. Figure 1-2 illustrates this interrelationship, often referred to as "adaptive management." In this continually cycling process, monitoring provides the information necessary to identify problems and to establish biocriteria for the decision making, management planning, and implementation necessary to respond appropriately. Continued monitoring then reveals the relative success of the effort by comparing the new results to those criteria again. At this point the criteria or the management plan may be adjusted as needed and the cycle repeats. Ideally, the estuarine or coastal waters improve with each cycle.

Figure 1-2
Program
Interdependence
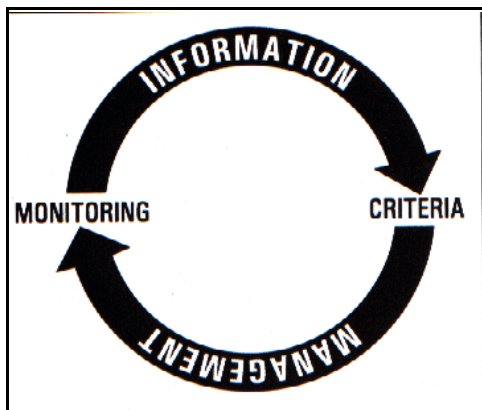
1.5 Implementing Biological Criteria

Implementing biocriteria requires an established and standardized methodology for biological assessment adjusted to regional or state conditions. Hence, guidance for state and regional development of biocriteria has two elements which are described in the biological criteria technical guidance documents such as this one:

▸ **Bioassessment Protocols** are methods used to assess the status and trends of water bodies. Guidance documents for bioassessment contain suggested methods and protocols for establishing monitoring programs that use biological assessment.

▸ **Biocriteria Guidance** assists states in establishing biological criteria for water bodies. Biocriteria are a series of ambient water resource quality values or statements of condition that relate to the desired biological integrity for that class of waters. When established they can be used to evaluate similar water bodies in that region. Implementation of biocriteria requires use of bioassessment protocols and a state or regional biomonitoring database. The National Program Guidance for biocriteria describes issues related to development and implementation (USEPA 1990). The first biocriteria technical guidance issued was for streams and small rivers (USEPA 1996a). It incorporated both biosurvey techniques and biocriteria development methods. It was followed by the Lakes and Reservoir Bioassessment and Biocriteria Guidance (USEPA 1998). Each of these documents incorporated biosurvey techniques and the same approach is being followed in similar documents for rivers, wetlands, and coral reefs in addition to this present technical guidance for estuaries and coastal marine waters.

1.6 Characteristics of Effective Biocriteria

Generally, effective biocriteria share several common characteristics:

▶ Provide for scientifically sound, cost-effective evaluations;

▶ Protect sensitive biological values;

▶ Protect healthy, natural aquatic communities;

▶ Support and strive for protection of chemical, physical, and biological integrity;

▶ May include specific characteristics required for attainment of designated use;

▶ Are clearly written and easily understood;

▶ Adhere to the philosophy and policy of nondegradation of water resource quality;

▶ Are defensible in a court of law.

In addition, effective biocriteria are set at levels sensitive to anthropogenic impacts; they are not set so high that sites that have reached their full potential are considered as failing to meet the criterion, nor so low that unacceptably impaired sites are rated as meeting them, which defeats the purpose of the CWA. The establishment of formal biocriteria warrants careful consideration of planning, management, and regulatory goals and the best attainable condition at a site. Balanced biocriteria will allow multiple uses to be considered so that any conflicting uses are evaluated at the outset. The best balance is achieved by developing biocriteria that closely represent the natural biota, protect against further degradation, and stimulate restoration of degraded sites.

Developing and implementing biological criteria occurs in three steps (USEPA 1996a):

1. Planning the biocriteria development program, including:

   • definition of program objectives;
   • establishment of interagency cooperation;
   • identifying acceptable levels of uncertainty for decisions made on the basis of biocriteria;
   • establishing data quality objectives.

2. Characterizing reference conditions for biocriteria and identifying candidate reference sites, which may require a biological survey.

3. Establishing biocriteria based, in part, on characterized reference conditions and designated use classes of the state.

1.7 Conceptual Framework

The central principle of biological assessment is comparison of the biological resources of a water body to a biological criterion based, in part, on a reference condition. Impairment of the water body is judged by its departure from the biocriteria. This approach presumes that the purpose of management is to prevent and repair anthropogenic; i.e., human-induced, damage to natural resources. Biological assessment of water bodies is predicated on our ability to define, measure, and compare biological integrity between similar systems. This requires an

operational definition of biological integrity as follows:

> "...the condition of the aquatic community inhabiting unimpaired water bodies of a specified habitat as measured by community structure and function (USEPA 1990)."

The functional definition also requires definitions of "unimpaired" and "community structure and function", and the habitat must be specified. Community structure and function is operationally defined by the biological measures chosen for bioassessment, consisting primarily of measures of species richness, trophic diversity (relative numbers of herbivores and top carnivores), and indicator species. In addition to biological community structure and function, chemical (DO, salinity, contaminants, dissolved TOC, inorganic nitrogen, etc.) and physical (sediment composition) attributes are measured to define an unimpaired site. The combined attributes form the basis for defining reference conditions for biological criteria. When unimpaired water bodies do not exist within a region, an operational definition of unimpaired can be developed from a combination of minimally impaired estuaries and coastal waters, historical information, and professional judgment (Section 1.7.2). Figure 1-3 shows a simplified framework for progressing from an estuarine classification to assessing the health of the estuary.

1.7.1 Indicators of Biological Integrity and Survey Protocols

Several analytical approaches have been developed to assess the biological condition of waterbodies within the framework of comparison to reference, ranging in complexity from simple comparison of indicator values, to

development of multivariate models:

▸ Comparison of indicator values — Indicator of metric values can be compared directly to the reference condition, without development of an index. This has been used most often for paleoecological comparison, where biological indicators are limited to certain indicator species, deposition rates, organic carbon loss, etc. (Turner and Rabalais 1994, Sen Gupta et al. 1996, Cooper and Brush 1991, Latimer et al. 1997).

▸ Multimetric index — The multimetric approach is to define an array of metrics or measures that individually provide limited information on biological status, but when integrated, function as an overall indicator of biological condition. Metrics incorporate information from individual, population, and community levels into a single, ecologically-based index of water resource quality (Gray 1989, Plafkin et al. 1989, Karr 1991). The index is typically a sum or an average of standardized scores of its component metrics (Barbour et al. 1999). Developed initially for streams, the multimetric approach has increasingly been applied to estuaries (Weisberg 1997, Hyland et al. 1998).

▸ Discriminant analysis to develop an index from metric values — In this approach, metrics (calculated as above) are used to develop a multivariate discriminant analysis model to distinguish reference sites from impaired sites. The calibrated model is then applied to assessment sites to determine whether they are impaired. This approach was used in EMAP-Near Coastal for the
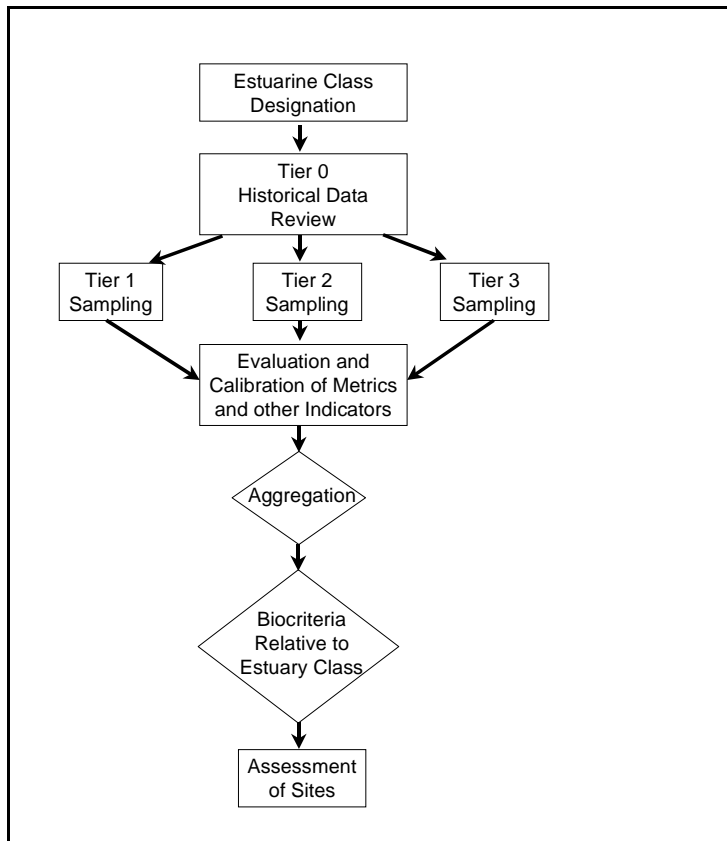
```
                    ┌─────────────────┐
                    │ Estuarine Class │
                    │   Designation   │
                    └─────────────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │     Tier 0      │
                    │ Historical Data │
                    │     Review      │
                    └─────────────────┘
              ┌──────────────┼──────────────┐
              ▼              ▼              ▼
        ┌──────────┐  ┌──────────┐  ┌──────────┐
        │  Tier 1  │  │  Tier 2  │  │  Tier 3  │
        │ Sampling │  │ Sampling │  │ Sampling │
        └──────────┘  └──────────┘  └──────────┘
              └──────────────┼──────────────┘
                             ▼
                    ┌───────────────────┐
                    │  Evaluation and   │
                    │ Calibration of Metrics │
                    │ and other Indicators │
                    └───────────────────┘
                             ▼
                        ◇ Aggregation ◇
                             ▼
                        ◇ Biocriteria ◇
                        Relative to
                        Estuary Class
                             ▼
                    ┌─────────────┐
                    │ Assessment  │
                    │  of Sites   │
                    └─────────────┘
```

Virginian and Gulf provinces (Paul et al. 1999, Engle et al. 1999).

▸ Multivariate ordination approaches — Several approaches have been developed using multivariate ordination to examine differences in species composition between reference and impaired sites. The purpose of ordination analysis is to reduce the complexity of many variables (for example, abundances of over 100 species from many estuarine sites), by re-ordering the information into fewer variables. These approaches have been used to show the effects of oil drilling in the North Sea (Warwick and Clarke 1991), and to develop an index of benthic quality in California (Smith et al. 2000).

While all of these approaches are appropriate to biocriteria development when properly applied, the multimetric approach is highlighted in this guidance. This is because it is the best developed and most extensively used method to date. Investigators should carefully consider what is most appropriate for their specific program. Time and experience will ultimately determine the best approach or combination for each state to use. Chapter 11 goes into further detail about methods of classification and assessment using all three approaches.

The multimetric concept came to fruition with the fish Index of Biotic Integrity (IBI) first conceived by Karr (1981). The IBI aggregates various elements and surrogate measures of process into a single assessment of biological condition. Karr (1981) and Karr et al. (1986) demonstrated that combinations of these attributes or metrics provide valuable synthetic assessments of the status of water resources.

A metric is a calculated term or enumeration representing some aspect of biological assemblage structure, function, or other measurable characteristic. Similarly, each of the assemblages (e.g., fish, benthic macroinvertebrates composing the aquatic community) measured would be expected to have a response range to perturbation events or degraded conditions. Thus, biosurveys targeting multiple species and assemblages; i.e., multimetric, will likely provide detection capability over a broad range of impacts, and the biocriteria derived from their results could provide protection to a large segment of the ecosystem.

Metrics can be expressed numerically as integers or ratios. Consistent routines in normalizing individual metric values provide a means of combining metric scores which initially consisted of dissimilar numerical expressions. However, final decisions on impact/no impact or management actions are not made on the single, aggregated value alone. Rather, if comparisons to established reference values indicate an impairment in biological condition, component parameters (or metrics) are examined for their individual effects on the aggregated value and for indications of potential causes.

Assessment of biological integrity using this multimetric approach typically focuses on four broad classes of community properties. Ecological systems respond to anthropogenic impacts with changes in one or more of these classes of properties (e.g., Karr et al. 1986, Schindler 1988, Plafkin et al. 1989, Schindler et al. 1989, Karr 1991, Barbour et al. 1992). The four properties are:

▸   **Health** of populations, typically expressed as number of individuals per m^2 or as biomass, reflecting possible stress from anthropogenic sources;

▸   **Community structure and composition**, or the number and kinds of species in an assemblage. Exotic species are typically undesirable, and high diversity is usually desirable. Species structure metrics include diversity and evenness indexes as well as presence of indicator species, counts of tolerant or intolerant species, the percentage of individual taxa in comparison to the total number sampled, and abundance proportions of taxonomic groups (e.g. crustaceans, mollusks, polychaetes), or comparisons of infauna vs. epifauna;

▸   **Trophic structure**, or the relative proportion of different trophic levels and functional feeding groups (e.g., Barbour et al. 1992). In estuaries, abundant, diverse, and relatively large top carnivores (e.g., piscivorous fish) are typically desirable as representative of a broad, stable, and substantial trophic network;

▸   **System function**, or the productivity and material cycling of the system or its components (trophic levels, assemblages, species). Measures of system function include primary production and standing stock biomass.

Since biological integrity is defined as an indicator of undisturbed conditions, it too must be measured relative to those conditions. The requirement of the biological criteria process for a reference by which to measure biological integrity makes it a practical tool (*sensu* Peters 1991) for managing society's impact on the natural environment.

Monitoring and assessment programs typically do not have the resources to measure all ecological attributes of concern to the public and to managers, and assessment tools must be cost-effective. Ideally, metrics selected for monitoring must be scientifically valid; should not require large amounts of expensive equipment; and should be relatively rapid in the field. The selected variables must be:

▸ **Related to Biological Integrity** In general, almost any biological measurement is related to biological integrity, but some are more clearly tied to the properties of biotic systems of concern to society (e.g., native species, fish production, diverse trophic structure) (Suter 1993);

▸ **Responsive to Environmental Stresses** Biological measurements and the metrics developed from them must respond to environmental stress. Metrics that are not monotonic; i.e., they do not consistently exhibit low values in response to one end of a stressor continuum and high values in response to the opposite end, or that respond oppositely to different stresses, are difficult to interpret in practice;

▸ **Measurable with Low Error** Variability and measurement error should be controllable so that a reasonable sampling effort yields sufficient precision. Index period sampling; i.e., sampling during specific time periods in the annual cycle, is one way to reduce seasonal variability. However, there are costs in terms of information derived which may be prohibitive (see later discussion on seasonality);

▸ **Cost-effective** Cost of a metric should be proportional to the value of the information obtained. Usually, the simplest approach is most cost-effective and should be selected so long as results are sufficient to the agency's objectives;

▸ **Environmentally Benign to Measure** Sampling methods that significantly disturb or alter habitats and biota should be avoided.

1.7.2 Comparison to a Reference

As noted earlier, establishing biocriteria includes determining the reference condition. The reference condition establishes the basis for making comparisons and for detecting use impairment. Because absolutely pristine estuarine and coastal marine habitats probably do not exist, resource managers must decide on acceptable levels of minimum impacts that exist or that are achievable in a given region. Acceptable reference conditions will differ among geographic regions and states because estuarine salinity gradients, trophic state, bottom sediment types, morphology and biological communities differ between regions.

Reference conditions can be established in a variety of ways. It is important to recognize that the reference condition is best developed from a population of sites, not from a single site. However, in some instances, particularly coastal environments and sites influenced by controversial land uses, the use of site-specific nearfield/farfield stations may be necessary and appropriate to augment the reference condition. They should include information derived from:

▸ **Historical Data** are usually available that describe biological conditions in

the estuary or coastal marine region over some period of time in the past. Careful review and evaluation of these data provide insight about the communities that once existed and/or those that may be reestablished. Review of the literature and existing data is an important initial phase in the biocriteria development process. However, if data have not been collected for this specific purpose, they need to be carefully reviewed before being applied;

▸ **Reference Sites** are minimally impaired locations in the same or similar water bodies and habitat types at which data are collected for comparison with test sites. Reference sites could include sites that are away from point sources or concentrated nonpoint loadings; sites in sub-estuaries; sites occurring along impact gradients (nearfield/farfield); and regional reference sites that may be applied to a variety of test sites in a given area;

▸ **Models** include mathematical models (logical constructs following from first principles and assumptions), statistical models (built from observed relationships between variables), or a combination of the two. Paleobiological reconstructions of historic or prehistoric conditions are typically statistical or empirical models (Latimer et al. 1997, Alve 1991, Dixit et al. 1992). The degree of complexity of mathematical models to predict reference conditions is potentially unlimited with attendant increased costs and loss of predictive ability as complexity increases (Peters 1991). Mathematical models that predict biological reference conditions should only be used with great caution, because they are

complex and often untestable hypotheses (Oreskes et al. 1994, Peters 1991);

▸ **Expert Opinion/Consensus** A consensus of qualified experts is always needed for assessing all of the above information; establishing the reference condition; and helping develop the biocriteria. This is especially the case in impaired locales where no candidate reference sites are acceptable and models are deemed unreliable. In these cases, expert consensus is a workable alternative used to establish reference "expectations". Under such circumstances, the reference condition may be defined using a consensus of expert opinion based on sound ecological principles applicable to a region of interest. The procedures for these determinations and decisions should be well documented for the record.

1.7.3 Assessment Tiers

Biological surveys of estuaries and coastal marine waters can be implemented in several tiers, ranging from a simple and inexpensive screening to detailed field sampling, analysis, and assessment. The tiered approach gives agencies one suggested approach for planning, organizing, and implementing biological surveys. Other approaches may also be available. Agencies should consider the approach that would work best to meet their program objectives. The tiers are intended to be implemented cumulatively, that is, each tier should incorporate the elements in the preceding tier as appropriate for the waters in which they are applied. Each integrated tier includes both biological and habitat components. Higher tiers require successively more effort and yield more detailed information on

specific biotic assemblages and potential stresses on the system. Higher tiers reflect higher quality information and reduced uncertainty in the final assessment (Costanza et al. 1992). A desktop screening and three field survey tiers are described in this document. Figure 1-4 provides a summary of the requirements for each tier.

Tier 0 is a desktop screening assessment that consists of compiling documented information for the estuary or coastal marine areas of concern through a literature search and sending survey questionnaires to local experts. No field observations are made at this assessment level. Desktop screening should precede any of the three subsequent tiers. Its purpose is to support the planning for monitoring and more detailed assessments. Information to be compiled in Tier 0 includes: area and geomorphometric classification, habitat type, watershed land use, population density, NPDES discharges, water quality data (salinity, temperature, DO, pH, turbidity), biological assemblage data, and water column and bottom characteristics.

Tier 1 is the least complex of the survey approaches. It consists of a one-time visit to sites during a suitable, predetermined index period to collect biological and habitat data using standardized methods. The focus of this tier is on developing screening or survey information. These variables include a rudimentary identification of organisms (benthos, fish, macrophytes, or phytoplankon), water column characteristics (salinity, temperature, DO, pH, Secchi depth, water depth), and bottom characteristics (grain size, RPD layer depth, total volatile solids, and sediment toxicity). States may choose some variation of this list depending on regional characteristics and resources. Evaluation of the data collected, as well

as historical data for the area, leads to an initial classification of sites and identification of candidate reference sites.

Tier 2 is somewhat more complex. A higher level of detail is incorporated into the standardized biological methods and multiple visits to the site are made to address temporal variability and/or seasonality. Another assemblage (epibenthos) could be selected in addition to those listed above.   Water column nutrient measurements are added to the Tier 1 water column characteristics. A tactile categorization of grain size, plus total organic carbon, are added to the bottom characteristics. The data collected in this tier will allow the development of preliminary biological criteria.

Tier 3 is the most rigorous survey tier. Three or more assemblages are sampled here, through multiple site visits to account for seasonal variations in the selected estuarine and coastal marine biological assemblages and should incorporate supplemental studies which might be necessary for diagnostic assessment of the potential causes of observed impairments. This tier adds water column pesticides and metals measurements, plus full grain size characterization (sieving to determine percent grain size composition), acid volatile sulfides, and sediment contaminants. This tier also allows the resource agency to develop a database sufficient to support resource management activities to reduce the identified impairments and to develop and refine biocriteria.

Biological Assessment

The procedure of biological assessment is to sample two or more biological assemblages and record data such as abundance, condition, biomass, and

Figure 1-4
General comparison of Tiered Approach. Tiers are intended to be implemented cumulatively. Each tier should incorporate the elements in the preceding tier as appropriate for the waters in which they are applied, as necessary for specific programs.

Tier 0
-No field observations
-Desktop screening
-Literature search
-Questionnaires to local experts
-Support planning for monitoring and more detailed assessments

---

Tier 1
-One time visit to sites during suitable, predetermined index period
-Least complex survey approach
-Develop screening/survey information
-States choose variation of variables (assemblages + water column & bottom characteristics) according to regional characteristics & resources
-Leads to initial classification & ID of candidate reference sites

---

Tier 2
-2 or more visits to site
-More complex
-Possible to add another assemblage
-Add to water column & bottom characteristics samples
-Allows for development of preliminary biological criteria

Tier 3
-4 or more visits to sites
-Most rigorous
-3 or more assemblages
-Incorporate supplemental studies
-Additions to water column & bottom characteristics
-Develop database to support resource management activities to reduce impairments & define/refine biocriteria

---

other characteristics of each species. These data are then used to calculate metrics, such as taxa richness, percent dominance, number of intolerant species, and percent abundance of tolerant species. Each metric is compared to its expected value under reference conditions, and rated good (similar to reference), fair (different from reference), or poor (substantially different from reference). Numeric scores are assigned to the ratings, and the scores of all metrics of an assemblage are summed for a total score for the assemblage. The total score is again compared to the expected total score under reference conditions, and the assemblage as a whole is assigned an ordinal rating of good, fair, or poor. This second comparison to reference conditions is necessary because not all metrics are expected to score "good" at all times even in pristine conditions; the final assemblage score thus takes into account natural variability in metric values.

Once these values are satisfactorily established they can be incorporated in the development of a biocriterion for a particular estuarine or coastal marine

class. "Biological assessment" at this point becomes a comparison of monitoring scores to the biocriteria for management decision making. The following several chapters describe the processes necessary to the development of suitable metrics and finally their incorporation in biological criteria for water resource management decision making.

Chapter 2
Biological Survey

2.1 Indicators of Biological Integrity

A key concept underlying the approach to biological surveys presented in this document is that of biological integrity. Biological integrity, discussed in greater detail in Section 1.7, may be operationally defined as

> "...the condition of the aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by community structure and function (USEPA 1990)."

Biological integrity is an ideal condition; estuarine and coastal marine communities can approach a condition of biological integrity when they are minimally impaired by human activities. In order to determine the degree to which these communities approach biological integrity, it is necessary to measure attributes (or indicators) of community structure and function and to be able to distinguish between natural variations and anthropogenic impacts.

Various techniques can be used at any level to document the effects of anthropogenic perturbations on biological communities. Discussion of these techniques falls into three general areas, the first two of which are measurement processes and the third is a data processing technique. They are:

- ▸ Measures of community condition and change;

- ▸ The presence or absence of indicator taxa;

- ▸ The use of indexes to compile and evaluate large amounts of biological data for evaluation.

The suitability of many of the approaches in each of these categories has long been the subject of debate among biologists and natural resource managers. The following discussion examines both the utility and uncertainty surrounding these community assessment tools.

2.2 Primary Measures of Community Condition and Change

Whenever possible, the investigator should try to examine two or more assemblages because different organism groups react differently to perturbation. The more diverse the measures used, the more robust the investigative technique is and the more confidence the manager can place in the results. However, this idea must be reconciled with the limitations of the costs of multiple and diverse surveys and the relative availability of reliable scientific methods to measure some assemblages. The prevalent approaches today are measures of benthic macroinvertebrate infauna, fish, and aquatic vegetation.

2.2.1 Benthic Macroinvertebrates

The benthic infauna have long been used for water quality assessments because of their tendency to be more sedentary and thus more reliable site indicators over time compared to fish and plankton. Consequently, a larger body of data has been accumulated for this assemblage. Examination of benthic

---

community structure and function is a valuable tool for evaluating the condition of benthic habitats, for monitoring rates of recovery after environmental perturbations and potentially to provide an early warning of developing impacts to the system. Bilyard (1987) and USEPA (1991) cite the following specific advantages of monitoring benthic infauna to determine overall aquatic community health:

▸ Benthic infauna are typically sedentary and therefore are most likely to respond to local environmental impacts, thus narrowing the list of possible causes of impairment;

▸ Benthic infauna are sensitive to disturbances of habitat such that the communities respond fairly quickly with changes in species composition and abundance;

▸ Benthic infauna are important components of the food chain and often act to transport not only nutrients, but also toxicants, to the rest of the system;

▸ Monitoring benthic infauna provides an *in situ* measure of relative biotic integrity and habitat quality;

▸ Of the biota typically measured, this assemblage has the strongest supporting database. Thus, it has extensive historical and geographic application.

Some limitations of benthic infauna sampling include:

▸ Relatively few state and federal programs have the necessary in-house taxonomic expertise to support extensive monitoring activities;

▸ Current methods can distinguish severely impaired sites from those that are minimally impaired. However, it can be difficult to discriminate between slightly or moderately impaired areas, particularly in estuaries (due to their natural spatial and temporal variability);

▸ The condition of benthic habitats can vary over relatively small scales. Therefore, if too few samples are collected from a specified area, the ambient heterogeneity to be expected may be missed, potentially leading to incorrect conclusions regarding the biological and water quality conditions in the area;

▸ The cost and effort to sort, count, and identify benthic invertebrate samples can be significant, requiring tradeoffs between expenses and the desired level of confidence in decisions based upon the collected data.

2.2.2 Fish

Fish are an important component of estuarine and marine communities because of their economic, recreational, aesthetic and ecological roles. The abundance and health of the fish community is also the primary indicator used by the public to discern the health of a water body. Fish are good indicators of ecological health because:

▸ They are relatively sensitive to most habitat disturbances;

▸ Being mobile, sensitive fish species may avoid stressful environments, leading to measurable population patterns reflecting that stress;

▸ Fish are important in the linkage between benthic and pelagic food webs;

▸ They are long-lived and are therefore good indicators of long-term effects;

▸ They may exhibit physiological, morphological, or behavioral responses to stresses;

▸ Fish may exhibit obvious external anatomical pathology due to chemical pollutants;

▸ Fish databases originally compiled to support state and federal fisheries management programs may be available. These databases may require integration with other data (e.g., water quality) to be useful for bioassessment and biocriteria purposes.

The limitations on the use of fish in community bioassessments include:

▸ Fish represent a relatively high trophic level, and lower level organisms may provide an earlier indication of water quality problems;

▸ Some fish are resident species with relatively limited lifetime spatial ranges. Others have relatively large ranges, making it difficult to isolate probable causes of degradation that could occur anywhere within their range. Thus, the spatial scale of sampling is an issue and because of seasonal, open water migrations, temporal adjustments may also be necessary;

▸ Mobile organisms such as fish may avoid stressful environments, reducing their exposure to toxic or other harmful conditions;

▸ Fish surveys may be biased because of recreational and commercial fishing pressures on the same or related fish assemblages;

▸ Some fish are very habitat selective and their habitats may not be easily sampled (e.g., reef- or marsh-dwelling species);

▸ Since they are mobile, spatial variability is very high, requiring a large sampling effort to adequately characterize the fish assemblage.

2.2.3 Aquatic Macrophytes

Aquatic macrophytes in estuarine and coastal marine waters may include vascular plants (e.g., seagrasses) and algae (e.g., sessile and drift). Vascular aquatic macrophytes are a vital resource because of their value as extensive primary producers in estuaries. They are a food source for waterfowl, a habitat and nursery area for commercially and recreationally important fish species, a protection against shoreline erosion, and a buffering mechanism for excessive nutrient loadings. The primary productivity that has been observed for submerged aquatic vegetation (SAV) communities in estuaries is among the highest for any aquatic system (USEPA 1992). Excessive nutrient loadings lead to prolific phytoplankton and epiphytic macroalgal growth on seagrass which out-compete the seagrass through shading, as evidenced by the 1970s and 1980s decline of eelgrass in the Chesapeake Bay along with the current decline in Waquoit Bay. Because of the combined high productivity and habitat function of this plant community, any or all of the other estuarine or coastal marine biota can be affected by the presence or absence of macrophytes.

Some of the advantages of using aquatic macrophytes in biological surveys are:

▸ Vascular plants are a sessile community. There is essentially no mobility to rooted vascular or holdfast-established algal plant communities, so expansion or contraction of seagrass beds can be readily measured as an environmental indicator;

▸ Measurement of macrophyte community extent and relative density can be fairly easily accomplished by remote means, such as aerial photography, if the water is clear or shallow;

▸ Sampling frequency is reduced because of the relatively low community turnover compared to other biota such as benthic invertebrates or fish;

▸ Taxonomic identification in a given area is generally consistent and straight-forward.

Some of the disadvantages of macrophyte surveys are:

▸ Relatively slow response by the plant community to perturbation makes this a delayed indicator of water quality impacts. This could be critical if prompt management responses are needed;

▸ Successional blooms of some macrophytes means seasonal cycles need to be identified and accommodated by the survey schedule to avoid misinterpretation of data and false assumptions of water quality impacts;

▸ Changes in abundance and extent of submerged macrophytes are not necessarily related to changes in water quality;

▸ Aquatic macrophytes do not stand alone as an indicator of ecosystem condition; additional parameters (e.g., water column nutrient concentrations, light penetration) are required to interpret macrophyte data.

2.2.4 Phytoplankton

Many estuaries and marine waters can be considered "plankton-dominated" systems, which implies that this assemblage should provide valuable information in an assessment of ecosystem condition. Advantages of using plankton include:

▸ Plankton provide the most notable indication of eutrophication in estuarine environments. Changes in nutrient concentrations can result in long-term changes in estuarine community structure and function and planktonic primary producers are one of the earliest communities to respond;

▸ Changes in plankton primary production will in turn affect higher trophic levels of macroinvertebrates and fish;

▸ Many states routinely monitor chlorophyll *a* as part of water quality monitoring due to the ease and relatively low cost of analysis;

▸ Plankton have generally short life cycles and rapid reproduction rates making them valuable indicators of short-term impact.

As with all other assemblages, there are disadvantages associated with using phytoplankton in a biosurvey:

Biological Survey

- The fact that phytoplankton are subject to rapid distribution with the winds, tides, and currents means they may not remain in place long enough to be source identifiers of short-term impacts. This problem is compounded by the ability of some phytoplankton to synthesize atmospheric sources of nitrogen, thus confounding the identification of runoff sources of nutrients in estuaries and the resultant changes in the aquatic biota;

- Taxonomic identification of phytoplankton can be difficult and time-consuming;

- Competition by aquatic macrophytes, higher respiration rates, and increased grazing by zooplankton may counteract increased phytoplankton biomass resulting from nutrient enrichment. These reasons argue for investigating phytoplankton and zooplankton together as biological indicators;

- Phytoplankton can undergo blooms, the causes of which might be indeterminate, at varying frequencies.

2.3 Measures of Community Condition and Change Being Developed

Two assemblages (zooplankton, epibenthos) have considerable potential for expanding the biological information available for biocriteria development and bioassessments. These assemblages, however, are considered "developmental" at this time. As survey methods become more refined and routine, databases for these assemblages will expand and the techniques are expected to become sufficiently robust to be incorporated in biocriteria development and environmental management decision making. Paleoenvironmental reconstruction is an additional technique being developed. This technique allows investigators to infer past conditions from the remains of several groups of organisms found in sediment cores, and to compare those past conditions to current ones.

2.3.1 Zooplankton

Zooplankton consist of two basic categories: holoplankton which spend their entire life cycle as plankton, and meroplankton which are only plankton while in the larval life stage.

Holoplankton are characterized by rapid growth rates, broad physiological tolerance ranges, and behavioral patterns which promote their survival in estuarine and marine waters. The calanoid copepods are the numerically dominant group of the holoplankton, and the genus *Acartia* (*A. tonsa* and *A. clausi*) is the most abundant and widespread in estuaries. *Acartia* is able to withstand fresh to hypersaline waters and temperatures ranging from $0°$ to $40°C$.

The meroplankton are much more diverse than the holoplankton and consist of the larvae of polychaetes, barnacles, mollusks, bryozoans, echinoderms, and tunicates as well as the eggs, larvae, and young of crustaceans and fish.

Zooplankton populations are subject to extensive seasonal fluctuations reflecting hydrologic processes, recruitment, food sources, temperature, and predation. They are of considerable importance as the link between planktonic primary producers and higher carnivores. As such, they are also early indicators of trophic shifts in the aquatic system.

Advantages of zooplankton sampling are similar to phytoplankton:

▸ The rapid turnover of the community provides a quick response indicator to water quality perturbation;

▸ Sampling equipment is inexpensive and easily used;

▸ Compared to phytoplankton, sorting and identification is fairly easy.

Some limitations of using zooplankton in biosurveys are:

▸ The lack of a substantial data base for most regions;

▸ The high mobility and turnover rate of zooplankton in the water column. While this permits a quick response by zooplankton to environmental changes on the one hand, it also increases the difficulty of evaluating cause and effect relationships for this assemblage.

2.3.2 Epibenthos

The sampling of those animals living on the sediments or on structures may prove to be the link between relatively low cost but highly variable fish community information, and the more consistent but expensive benthic macroinvertebrate surveys. The process has been tested with considerable success in Washington, North Carolina, and Florida (Chapter 13).

Advantages of using this assemblage are:

▸ The relatively sedentary life style of some epibenthic fauna can result in an in-place accumulation of indicative pathogens and toxicants in individuals while the community composition reflects the average salinity, temperature and dissolved oxygen of that locale over an extended period of time (Day et al. 1989);

▸ Ease of data collection by use of small otter trawls or beam trawls;

▸ Relative ease of identification because taxonomic lists of local crustaceans, mollusks, and echinoderms can be fairly easily compiled;

▸ Sampling is as inexpensive as fish surveys, and can often be done with the same or similar equipment during the same survey;

▸ Decapod crustacea are usually very important prey for fish and are important components in benthic food webs. Some (e.g., shrimp and crabs) are harvested for human consumption.

Possible difficulties involve:

▸ Potential equipment snags and difficulties in macrophyte beds;

▸ Benthic infauna would likely be included in the trawl sample due to disturbance of surface sediments;

▸ As when using otter trawls for fish, benthic habitat may be destroyed;

▸ There is greater potential for avoidance by organisms than when sampling for benthic macroinvertebrates, though not as great as with fish surveys;

▸ Because of relatively low taxa numbers in some environments, especially coastal marine waters, impact response may not be as sensitive as desired; this could be

addressed by the use of indicator species instead of a multimetric approach;

▸ Epibenthos are very sensitive to substrate type;

▸ Relative sensitivity remains to be determined in many areas.

2.3.3 Paleoenvironmental Reconstruction: preserved remains

Several groups of organisms in estuaries leave remains in the bottom sediments. Some of the remains are resistant to decay and become a permanent biological record of the life in that waterbody. Comparisons of present-day biota to that of the past allow past environmental conditions to be inferred. Several groups of organisms have been used for this type of study in estuaries including diatoms, dinoflagellates, and foraminifera (Latimer et al. 1997).

The approach is to elucidate relationships between environmental conditions (for example, temperature, dissolved oxygen, nutrient concentrations) and the relative abundance of target species. These known relationships are then used to infer past conditions from the observed remains in the sediment. Advantages of studying paleoenvironmental systems include:

▸ diatoms, dinoflagellate cysts, and foraminifera found in sediments integrate conditions over broad spatial scales and over time periods of one year or more, so that short-term variability does not confound assessment;

▸ there is no need to adhere to an index period for sampling;

▸ paleoenvironmental reconstruction can provide a site-specific reference by showing conditions in the past.

Disadvantages of studying paleoenvironmental systems include:

▸ it requires a relatively stable depositional environment; it is not suitable for shallow estuaries subject to frequent resuspension;

▸ it requires conditions for preservation of target assemblages in the sediment;

▸ temporal resolution is limited by the rate of accumulation (between 1-10 years); it cannot be used to assess short-term response to stressors or to restoration efforts;

▸ at the time of this writing, technical expertise for estuarine paleoecology is specialized, with only a small handful of research institutions active in North America.

2.4 The Use of Indexes to Compile And Evaluate Biological Data

It is evident that biological surveys can generate tremendous amounts of raw data. The usual approach to sorting this wealth of observations is to summarize a series of diverse community measurements into one or more dimensionless indexes, much as the cumulative performance of a student's work for a year can be reduced to annual grades.

As with student grades, the use of dimensionless indexes is a well-established and consistent way to evaluate and compare many discrete units as a continuum of performance or condition. Also similar to student

grading, detailed insight is lost when the complex interplay of so many discrete variables is reduced to a single score. The reasons for high or low scores are not always evident and the accuracy of the scoring process itself is always subject to debate. Indexing is the only way to rank order information for decision making. However, valuable insight is lost at every level of data reduction. There is no alternative to the process short of relying entirely on the professional judgment and wide variation of skill of individual biologists. The strengths of index development and use are:

▸ It is a rational, consistent way to reduce large amounts of data to unitless, meaningful interpretations;

▸ It is a quantitative treatment of the observations which permits statistical assessments;

▸ Interpretive bias is reduced in the treatment of the data.

Conversely, indexing:

▸ Removes the decision-making from detailed evaluation of the data and information to just reporting of simplified indexes;

▸ May be viewed as irrefutable, despite evidence to the contrary;

▸ May obscure important and confounding interrelationships in the aquatic environment contributing to the index score(s);

▸ Obscures more information as each level of data reduction is performed leading to an index value, so that some indexes are not sufficiently sensitive to reflect biotic change;

▸ Provides no indications of causes of the relative condition of the system.

The best way to guard against the problems of indexing, while using it to expedite decision-making, is to **always** retain the raw data. These files can be used to translate historical data sets into present indexes for temporal continuity, and even more important, they can be evaluated to provide an interpretation and potential diagnosis for management action when a particular site is being evaluated.

Indexes are most often used to measure community composition such as species abundance, diversity, evenness, richness, and dominance or conditions such as incidence of disease, malformation, and distributions of year classes. These can be used to assess the changes in community structure that occur as a result of anthropogenic perturbations (Boyle et al. 1990). Community function can also be described through indexes such as the Infaunal Trophic Index (Word 1978, 1980, USEPA 1987).

Although indexes have long been used in applied and theoretical ecology, it is recognized that some of them, when applied individually, are insensitive to stress-induced changes in naturally occurring biological communities (Boyle et al. 1990). Because of varying sensitivities of the community indexes, several of them should be used concurrently for evaluating impacts. This approach provides greater certainty of the data interpretation than reliance on any single index. Conversely, while Ludwig and Reynolds (1988) indicate that the most reliable community measures in evenly matched surveys are number of individuals and number of taxa as direct measures; it has been

observed in the coastal marine studies associated with this guidance manual that, at least in two mid-Atlantic Bight outfall studies, the diversity index and the richness index both appear to be more responsive than number of individuals or number of taxa to sewage impacts (Gibson, Chapter 13). For a more detailed discussion of the different indexes and their particular applications see Chapter 11 (Index Development) and Chapter 13 (Case Studies).

2.5 Indicator Taxa

Indicator taxa or species are those organisms whose presence (or absence) at a site indicates specific environmental conditions. If an organism known to be intolerant of pollution is found to be abundant at a site, high water quality conditions can be inferred. On the other hand, dominance by pollution tolerant organisms implies a degraded condition. When available, indicator taxa are an important, cost-effective preliminary survey tool for site assessments. However, the investigator should always ascertain that absence of an indicator organism is a fact and not merely a reflection of insufficient sampling.

Swartz et al. (1985, 1986, 1994) have demonstrated the sensitivity of the amphipod *Rhepoxynius abronius* to the complex contaminant mixture that often characterizes coastal marine benthic pollution. Their studies were performed along pollution gradients from the Los Angeles County Sanitation Districts' sewage outfalls to control conditions in Santa Monica Bay. The results showed that there were significant increases in the concentration of most sediment contaminants and significant decreases in benthic taxa richness and abundance at stations where sediment was acutely

toxic to *R. abronius* (Swartz et al. 1985). More studies performed by Swartz et al. (1994) at a designated Superfund site in San Francisco Bay also showed that acute sediment toxicity lab tests of *R. abronius* correlated with biologically adverse sediment contamination in the field. Other EMAP studies (Summers et al. 1992) included a 10-day acute test using the tube-dwelling amphipod, *Ampelisca abdita*. The majority of sediments proving significantly toxic to *A. abdita* were found in Louisiana and Alabama estuarine waters.

A well-known indicator for degraded systems is the polychaete *Capitella capitata*. *C. capitata* and its related species are collectively known as the *C. capitata* complex. In general, the presence of this indicator species corresponds to a dominance of deposit feeders that colonize an area as organic pollution increases. Swartz et al. (1985) observed dominance of *Capitella* near sewage outfalls. A recent study in the Mid-Atlantic Bight by the U.S. Army Corps of Engineers (1996) suggests that the polychaete *Amastigos caperatus* may have indicator potential similar to the *Capitella* complex.

A problem with using pollution tolerant indicator organisms is that some of these organisms may be ubiquitous and found in naturally occurring organically enriched habitats as well as in minimally impaired waters. To be useful as an indicator, they must have displaced other, less robust taxa and have achieved numeric dominance. Tolerant and ubiquitous organisms can be found in sediments far away from sources of sewage pollution and long after plumes have dispersed.

The use of the concept of "clean" indicator species is less subject to this form of misinterpretation. These "clean"

or highly sensitive organisms are less likely to be found in both polluted and high quality habitats.

The best option may be the paired use of both pollution tolerant *and* intolerant indicator species. If both indicators change concurrently in opposite directions, more confidence can be placed in the interpretation.

As part of the biological survey process, individual indicator species are useful in reducing analytical costs. They are not only a valuable preliminary assessment tool, they are a cost-effective way to define the magnitude, spatial, and temporal extent of an impact (USEPA 1992). Selected indicators should possess the following characteristics (Green 1984):

▸   Provide sufficiently precise and accurate appraisals of:

— species of concern
— magnitude of anthropogenic disturbance;

▸   Be cost-effective and statistically reliable as an alternative to monitoring all critical community measures;

▸   Appropriate to the spatial and temporal scale demanded by the study objectives.

When indicator species are employed in tandem for impact investigations, a gradient of species distribution can often be identified. Such a gradient might progress from the most degraded waters, having low diversity communities dominated by pollution tolerant opportunistic species, to unimpaired or minimally impaired waters having diverse communities that are comprised of a wide range of taxa,

including pollution sensitive ones and some that are pollution tolerant.

Chapter 3
Habitat Characterization

Biological assessment in estuaries and coastal marine waters is built around assessing two separate ecosystem components: the habitat and the biota. The biota are the resident plant and animal assemblages in the water body. The condition of the biota depends in part on the quality of the physical-chemical environment of estuaries and coastal marine waters; i.e., habitat. Habitat is in turn influenced by natural and catastrophic events, climate, and human activities. These include:

▸ Seasonal variations in precipitation, temperature shifts, and wind or wave patterns;

▸ Introduced or extirpated species able to influence the habitat (such as burrowing organisms, plants, or diseases/parasites;

▸ Shifts in sedimentation or scouring patterns;

▸ Dredging and filling;

▸ Shoreline or basin construction;

▸ Bulkheading and jetty construction;

▸ A variety of land use and navigational practices.

This conceptual model is based on an understanding of the causal mechanisms of natural and anthropogenic stress effects in estuarine and coastal marine ecosystems.

Estuaries integrate processes because they receive and retain matter and energy released in the watershed. Many human activities directly affect aquatic habitat including discharges, agriculture, and urban land use which contribute materials (sediment, nutrients, contaminants) to the water body. Biological communities are directly affected by their physical habitat and water quality conditions, and also by direct human activities such as stocking or harvesting.

Components of biota are the biological assemblages, such as algae, aquatic macrophytes, benthos, epibenthos, plankton, and fish. Habitat components for the biological assessment of these assemblages are hierarchical, and include the watershed, the nearshore zone, the water column, and the sediment. An integrated assessment evaluates the condition of estuaries and coastal marine waters by aggregating data on components of both habitat and biota. The habitat component may be damaged by physical stress or chemical degradation from pollution. Thus, habitat studies may help identify causes of biological decline as well as being the important determinant of the types of biotic communities to be expected. This classification function is crucial to proper biocriteria development.

Habitat characterization is essential to the proper classification of sites. Although estuaries are by definition transitional zones between fresh water and the sea, and both estuaries and coastal marine waters incorporate many environmental gradients (e.g., salinity, sediment grain size, depth), individual locations and conditions are often defined categorically. Thus, a site may be characterized as oligohaline or mesohaline with respect to salinity, or sand or mud with respect to sediment

grain size. The composition of biological assemblages can vary dramatically along these habitat gradients, and valid comparisons of estuarine and coastal marine biological assemblages require that their habitats be correctly classified.

3.1   Flow and Hydrography

The type of estuarine ecosystem in a specific area is primarily controlled by the physical environment; i.e., geomorphology, climate, salinity, and the availability of fresh water. The absolute values of the abiotic factors are not as important as the degree of fluctuation of factors such as the microclimate, water movement, chemical cycling, and physical structure (Day et al. 1989). In addition, the residence time of water in an estuary can influence overall water column pollutant concentrations.

The abiotic features thought to be important in determining the specific nature of estuaries as proposed by Day et al. (1989) are:

▸ The degree of protection and buffering from direct oceanic forces;

▸ The quantity of freshwater input and associated dissolved and suspended materials;

▸ The water circulation patterns that are determined by riverine and tidal currents, geomorphology and wind. Tides play a critical role in influencing circulation, and biochemical and biological processes. In many coastal regimes, the wind-driven currents are more predominant than tidal and geomorphologically-induced currents;

▸ Depth—stronger interactions between the water column and the bottom occur in shallow estuaries, thereby expediting the release of sediment nutrients for use by the phytoplankton;

▸ Variability of salinity and the sharpness and pattern of the salinity gradient from the mouth to the headwaters. Water circulation influences the salinity gradient and the distribution of biological assemblages;

▸ The rate of geomorphological change generated by various physical forces that control sediment transport within the estuary.

These controlling abiotic features are discussed in more detail in the following sections.

3.1.1   Circulation and Tidal Regime

Circulation is the physical process that influences or controls many of the ecological processes occurring in an estuary, including the degree to which an estuary is dominated or modified by hydrodynamics. The three major driving forces behind the circulation patterns in estuaries are: (1) gravitational circulation; (2) tidal circulation; and (3) wind-driven circulation. Geostrophic forcing; i.e., the Coriolis effect, can significantly alter estuarine circulation patterns as does bathymetry. A notable example is the Chesapeake Bay, where lower salinities extend further south along the Bay's western shore in comparison to its eastern shore.

Gravitational circulation is induced by water masses of differing densities and the layering of fresh water inflow on top of more saline waters. These density differences cause the lighter fresh water

to flow over top of the saltwater to form what is commonly termed the "salt wedge". In what is termed the "classical" estuarine gravitational circulation, the pressure surfaces of the fresh water are tilted seaward and the pressure surfaces of the saltwater are tilted toward the head of the estuary. The shear stress forces that occur at the interface of these two water masses cause vertical mixing and an eventual equalizing of the pressure surfaces somewhere near mid-depth. The fresh water surface layer has a net seaward movement, the saltwater layer on the bottom has a net movement upward into the estuary and the interface is a zone of no net movement.

Typically, density differences and the resulting circulation are determined by salinity and the circulation as is described above. However, in arctic or subarctic estuaries, the fresh water inflow may be substantially colder than the saltwater causing the fresh water inflow to sink below the saltwater, reversing the expected circulation pattern. In shallow lagoons that receive little or no fresh water input such as Florida Bay or Laguna Madre, the evaporation rate can cause the salinity of the water within the lagoon to rise higher than the ocean waters. When this happens, the ocean water flows into the estuary on the surface and the estuarine water flows out on the bottom, resulting in reverse circulation. Most often, though, shallow lagoons are well-mixed by the wind and reverse circulation is only observed at the mouth of the lagoon.

Tidal circulation occurs when the ebb and flow of the tides becomes the driving force behind circulation. This is known to occur in estuaries with steep constrictions or with shallow depths and large tidal ranges (e.g., Puget Sound), and in the absence of density gradients or wind stress. In many estuarine systems, gravitational and tidal effects coexist.

Wind circulation is common in estuaries with open water, shallow water depths, a small tidal range, and low fresh water input. Although the effects of wind may be overshadowed by gravitational or tidal forces, periods of sustained wind can have dramatic ecological effects. In the Ten Thousand Islands region of the Florida Everglades, sustained northern winds virtually drain the water out of large portions of the estuary for extended periods of time, exposing large areas of mudflats. Winds that blow into an estuary can create a net flood flow and result in the inundation of marshes and grass lands. In lagoons as well, water can pile up on one side of the lagoon creating a seiche or a sloshing of the water back and forth within the lagoon.

Attention to these diverse, often variable, and always significant circulation patterns is essential to understanding the biotic distribution of coastal and estuarine environments. This must always be addressed before attempting to attribute such distribution to anthropogenic influences.

3.2 Habitat Types

Use of a single habitat to characterize biological assemblage condition would minimize the requirements for expenditure of time and resources. However, estuaries and coastal marine waters are inherently heterogeneous systems. By definition, a salinity gradient is present in any estuary. Greater numbers of species are typically observed at the marine end of the salinity gradient, with the fewest numbers observed in about 3-7-ppt (Weisberg et al. 1993). This may be due to the unpredictable nature with which

the brackish water zone varies along the length of the estuary; driven by the strength and intervals of freshwater inflow.

In addition to a salinity gradient, estuarine habitats vary in bottom substrates, and in the predominance of erosional versus depositional environments. The variations in these characteristics will lead to differences in the way pollutants and other stressors will effect the biota. For example, depositional environments occur where large amounts of terrigenous sediments are transported by rivers to embayments and where the water is sufficiently quiescent that fine-grained sediments settle out. Metals, synthetic organic pollutants, and other contaminants adsorb to fine-grained sediments (Holland 1990) and low-density organic detritus. Thus, the prevalence of depositional areas may reduce the likelihood of water column exposure of estuarine organisms to toxic materials, but may increase the exposure to burrowing organisms. Conversely, the water column of erosional zones is often highly enriched as resuspended phosphorus is episodically mobilized.

Habitats in estuaries and coastal marine waters can be classified into nine major categories. These habitats are summarized below in a progression from open, deep waters to decreasing depths near the shoreline. The choice to sample one or more of these nine habitats will be dictated by their areal extent and the nature of the problems being addressed.

3.2.1   Open Water

Sampling in open water may demonstrate phytoplankton blooms which might be symptomatic of eutrophication from anthropogenic inputs of phosphorus or nitrogen. It should be noted that not all phytoplankton blooms (e.g., red tides) result from anthropogenic stresses. Increases in the standing stocks of bacteria associated with fecal material can be used to identify the presence of sewage effluent. Open water (plankton and nekton) studies also allow assessment of pelagic food webs. Sampling of pelagic finfish also occurs in open estuarine and marine waters. Limitations of sampling in open water include a high degree of patchiness in the plankton and finfish assemblages, which necessitates the sampling of large volumes and areas of water before results can be described with acceptable precision. Because of the transitory residence time of water parcels moving through estuaries and the short life cycles of planktonic flora and fauna, a relatively high sampling frequency is necessary to distinguish signals from noise in this area.

3.2.2   Soft Bottom Substrates

A "soft bottom" deposit may be dominated by mud or fine- to relatively coarse-grained, hard-packed sand. All of these sediments can be sampled with appropriate grabs. Soft bottom substrates provide habitat for economically valuable clams, shrimp, and juvenile flatfishes. Muds have a high surface area-to-volume ratio, providing a large surface area for the adsorption of metals and organic pollutants. Also, fine-grained deposits are often rich in biogenic adhesives (mucopolysaccharide secreted by microbes and meio- and macrofauna), to which organic pollutants may adhere. Under anoxic conditions, these deposits are often called "black ma yonnaise." Conversely, sands have a lower surface area-to-volume ratio. Thus, there is less surface area available in the deposit to which pollutants may adsorb.

A limitation of sampling in soft substrates is that the samples must be sieved before individuals can be counted and identified. With medium or coarse sands this also means large volumes of sediment are retained on the sieves for subsequent processing. Thus, measurements cannot be readily made in the field. This is especially true for samples in which an abundance of organic debris (wood chips, seagrass leaves) masks the presence of small, cryptic organisms.

3.2.3 Hard Bottom Substrates

Hard bottom substrates can include offshore rocky outcrops; oyster, relic shell, and worm tube reefs; and relic limestone and coral outcrops. Oyster and mussel beds are major habitats supporting a complex community of various species essential to the biological diversity of estuaries such as Chesapeake, Florida, and Tampa Bays. In fjords the predominant subtidal habitat may be the rocky walls extending below the water line. An advantage of sampling these areas is that the macroalgae growing along these walls may be relatively sensitive to stress and can be used in bioassays or tested for bioaccumulation (Levin and Kimball 1984).

Hard bottom substrates frequently occur in high-energy environments common in portions of Alaska, Washington, Oregon, and northern California, as well as New England. For example, gravel/cobble beaches are typically subject to high incident wave energy. Accumulations of fine-grained sediments or organic detritus are not expected on hard substrates either because there is no riverine source or because the energy of the environment prevents their accumulation. In addition, subtidal rocky bottoms are difficult to sample remotely. A grab sampler cannot be used, for example, as it can in soft-bottom habitats. Typically, divers conduct transect observations and take photos.

Another limitation of sampling hard substrate is that the larvae and spores of rocky substrate organisms are typically planktonic; therefore, recruitment may be strongly influenced by factors outside the estuary. Sampling and analysis methods must be appropriate both to the specific type of hard substrate as well as regional characteristics due to the fact that hard bottom substrates can exist from rocky, high energy regions in California to carbonate sediment, low energy regions in Florida.

3.2.4 Aquatic Macrophytes

Macrophyte beds are among the most important estuarine habitats, both ecologically and economically. Seagrasses (in this text, used to describe macrophytes existing from tidal fresh to marine conditions) create an environment with a high degree of structural complexity. Ecological niches exist within the sediment, between the rhizomes, along the surfaces of the leaves, and in the protected portion of the water column within the bed. Thus, a seagrass bed is capable of supporting a highly diverse fauna. Seagrass beds are among the most productive aquatic habitats. In western Florida, *Thalassia testudinum* beds produce 8,100-gm^{-2} dry weight of leaves at maximum standing crop (Hillman et al. 1989). Seagrass beds provide "nursery areas" for juvenile fishes and decapod crustaceans, including many of economic importance. Seagrass leaves are known to bioaccumulate and possibly to bioconcentrate metals (Ward 1989).

The right mix of nutrients and water clarity are key in seagrass growth and survival in estuarine systems. A study

by Stevenson et al. (1993) demonstrated regrowth of SAV in the Choptank River of the Chesapeake Bay (mesohaline salinity regime) was associated with mean DIN <10-μM, mean DIP <0.35-μM; mean SPM (suspended particulate matter) <20-mgL^{-1}; mean chlorophyll *a* (in water column) <15-μgL^{-1}; and mean light attenuation coefficient (Kd)<2-m^{-1}.

In addition to serving as habitat for invertebrates and fish, macrophytes are also a biological assemblage in their own right and appear to be relatively sensitive to stress. These beds can be monitored from historical photos and other records (Shepard et al. 1989). Areal reductions are often attributed to shading, which may result from a variety of anthropogenic factors including turbidity (from ship traffic, dredging, or harbor construction which may stir fine-grained material up into the water column) and eutrophication (where inputs of nutrients stimulate increases in the density of phytoplankton and of epiphytic macro- and microscopic algae on leaf surfaces).

A limitation of sampling in seagrass habitats is that because of their ecological complexity, multiple sampling strategies are required to survey the various components of the fauna (Howard et al. 1989), increasing the expenditure of time and resources by investigating agencies.

3.2.5   Beaches

Beaches are accumulations of unconsolidated sediment (e.g., sand, cobble) extending shoreward from the near low-tide line to some demarcation such as a sea cliff or dune field, or to the point where permanent vegetation is established (Komar 1976). High energy beaches typically consist of shingle or cobble due to the removal of finer sediments by high incident wave energy. These beaches may be commonly found on the Atlantic coast north of Cape Cod and along the Pacific coast. High energy beaches of southeastern Florida are composed of sand and are protected by sabellariid reefs offshore. Low energy beaches consist of sand and finer sediments and are widely located on the Atlantic coast south of Cape Cod, the Gulf of Mexico, and along the southern Pacific coast, and along the shorelines of most estuaries and their tidal tributaries.

Distinct zones occur across the beach profile, proceeding from the subtidal offshore and inshore zones, through the foreshore that lies between the upper current of wave swash at high tide and the low water mark of the backrush of the wave swash at low tide, to the subaerial backshore (Komar 1976). These zones contain a gradation of faunal communities in response to varying conditions of wave energy, sediment size, and inundation by water. Sampling across the beach profile can be problematic, particularly on high energy beaches, due to the rapidly-varying wave conditions and distribution patterns of beach infauna and epifauna. If attempted, core samplers can be used on low energy beaches but quadrat surface sampling may be required on high energy beaches.

3.2.6   Sandflats

Deposits described as "fine sand" contain some silts and clays and harbor rich communities of both deposit- and suspension-feeding invertebrates. "Coarse sands," found in higher energy environments, are expected to be dominated by suspension-feeding invertebrates. If a sufficient erosional force is applied to a sandy bed by the overlying flow, sand waves form. Sand waves indicate a physically stressful habitat, typified by a relatively

depauperate fauna. It should be noted that coarse sands generally contain relatively low faunal densities and biomass. However, when coarse sands are colonized by surf clams and echinoderms (sand dollars), densities and biomass can be very high. Compared to mudflats, relatively little organic material and few associated pollutants are expected to accumulate in coarse sands. Calcareous deposits in Puerto Rico and Hawaii may provide an exception to this in that they have high organic matter content and many pollutants (Dossis and Warren 1981).

3.2.7   Mudflats

Mudflats provide habitat for commercially important clams and other invertebrates and feeding habitat for fishes and shorebirds. Fine-grained sediments (i.e., clay) and organic detritus accumulate on mudflats. These materials provide a large surface-to-volume ratio adsorb metals and organic pollutants. This same expanse of usually nutrient and oxygen rich habitat also supports diverse primary producers, bacteria, plankton, and fish and invertebrate species in a complex community.

3.2.8   Emergent Marshes

Where present, emergent marshes may be extremely important components of estuaries and coastal marine embayments and lagoons because they filter storm water runoff from cities, forests, and agricultural areas. Emergent marshes may also provide flood protection by attenuating peaks in storm water flow and reducing the erosive energy of wave action.

Traditionally, the ecological health of a marsh has been characterized by measuring changes in its areal extent and the abundance and diversity of its

fish and wildlife populations. Although "microelements" of the marsh community (e.g., terrestrial insect assemblages) may be more sensitive to perturbations, these studies may be more expensive and time-consuming to carry out. For example, insect populations are difficult to sample quantitatively because of their mobility and cyclic abundance.

3.2.9 Mangrove Forests

Low-lying tropical coasts are often bordered by dense mangrove forests. Temperature, particularly the frequency of freezes, and rainfall gradients restrict mangrove distribution. In the Gulf of Mexico, mangrove forests occur along the coasts south of Cedar Keys, Florida and generally south of Port Isabel, Texas. Isolated stands of black mangrove (*Avicennia germinans*) occur in the Mississippi deltaic plain (Day et al. 1989), but mangroves are otherwise absent from the northern Gulf coast, which is dominated by salt marshes. Significant mangrove stands also occur south of Cape Canaveral on the east coast of Florida (Odum and McIvor 1990). Mangrove forests and their associated waters provide valuable habitat for a range of invertebrates, fishes, amphibians, reptiles, birds, and mammals. Mangroves are also valuable as stabilizers of intertidal sediments, and the structural complexity of the prop-roots provides habitat for many commercially and recreationally important fishes.

3.3   Water Column Characteristics

Many chemical and biological processes in the environment are affected—directly or indirectly—by the physical characteristics of that environment (Thomann and Mueller 1987). Therefore, collection of physicochemical

data such as salinity, temperature, dissolved oxygen, pH, turbidity, nutrients, contaminants, and depth provides information necessary to evaluate biological data.

Although organisms living in estuaries are adapted for life in a physically dynamic system, many are living near the limit of their physiological tolerance range and any long-term alteration of physicochemical environmental conditions could force their permanent exclusion from the estuary. Even in minimally-impaired estuaries, causes of mass mortalities have been attributed to depletions of dissolved oxygen and changes in temperature, salinity, and excessive turbidity (Odum 1970). In areas subjected to anthropogenic stress, changes in physical and chemical parameters may occur too frequently, be increased in magnitude, or be sustained for periods of time that only the extremely tolerant organisms can endure.

The community composition of marine and estuarine fish assemblages is determined by the various species tolerances and preferences for environmental variables such as substrate, salinity, and temperature (Weinstein et al. 1980). These environmental variables are controlled by the quantity and direction of the flow of water; i.e., fresh water inflow, tidal cycles, circulation patterns, etc. The spatial and temporal distribution and abundance of fishes would typically be determined by these variables. Anthropogenic stress adds to the complexity by interfering with some aspects of the physiology of the fishes.

In the Chesapeake Bay, salinity is the major factor affecting the regional distribution of macrobenthos, while sediment characteristics have the most influence over local distributions (Carriker 1967). In the upper Chesapeake Bay, sedimentation, stratification, circulation, nutrient levels, and dissolved oxygen concentrations are all determined by the strength and variation of the fresh water inflow from the Susquehanna River.

Basic water quality parameters should always be monitored to provide a record of environmental conditions at the time of sampling and to provide information used in assessing the condition of biological assemblages. These parameters should be measured at the same time and location as the biological sampling. Such episodic data will only serve to provide a snapshot of the conditions at the time of sampling and will not characterize the habitat conditions in such dynamic ecosystems. To properly characterize many water quality conditions, long-term data sets are required, including data collected at short intervals over complete tidal cycles for each season.

Monitoring schemes for physicochemical water quality characteristics of the habitat usually involve *in situ* methods. Data on the water column's characteristics can be collected relatively inexpensively; the expense of different methods is generally governed by the level of automation. Nonetheless, it is essential to standardize the monitoring design in order to ensure the comparability of monitoring data throughout the program.

Measurements of temperature, salinity, and turbidity should be taken at a minimum of four depths in the vertical profile: (1) 1-m below the surface, (2) 1-m above the bottom, (3) 1-m above the pycnocline, and (4) 1-m below the pycnocline. If the waters are too shallow or no stratification occurs, it would be appropriate to take samples

just below the surface, at mid-depth, and just above the bottom. However, features of water masses recorded in the historical data; i.e., historical profiles of salinity, temperature, and turbidity, and the collection of other data types (e.g., plankton community structure and water chemistry) should be considered when establishing sample depths (Pond and Pickard 1983).

When using *in situ* methods (e.g., a conductivity-temperature-depth meter [CTD], temperature, salinity and pH measurements should be taken at 1-m intervals with a maximum interval of $\leq$5-m in deeper coastal waters. The measurements should be made over the entire depth profile (to within 1-m of the surface and bottom). Little additional cost is incurred for this detailed characterization of the water column once the CTD is deployed. In areas of high stratification, a smaller interval would be appropriate. With some of the newer, more expensive probes, a continuous readout is possible and discussion of depth intervals is immaterial.

3.3.1 Salinity

Estuaries are transitional zones in which the chemical composition varies from that of freshwater to marine. Salinity is a key determinant in the distribution of estuarine flora and fauna, especially for benthic invertebrate communities (e.g., Engle et al. 1994, Holland et al. 1987, Summers et al. 1993, Weisberg et al. 1993). Taxa richness is most strongly affected by salinity, with relatively low richness in brackish waters compared to freshwater and seawater. Taxa richness metrics can be expressed as a "percent of expected" for a given salinity (Engle et al. 1994, Summers et al. 1993, Weisberg et al. 1993).

The best-known estuarine zonation system (the Venice system) is based on salinity and establishes five estuarine salinity zones:

- ‣ Limnetic (0-0.5-ppt)
- ‣ Oligohaline (0.5-5-ppt)
- ‣ Mesohaline (5-18-ppt)
- ‣ Polyhaline (18-30-ppt)
- ‣ Euhaline (>30-ppt).

Bulger et al. (1993) conducted a Principal Components Analysis (PCA) to derive estuarine salinity zones based on field data on the salinity ranges of 316 species or life stages in the mid-Atlantic region (primarily species found in the Chesapeake and Delaware Bays). The PCA showed that the data structure underlying the salinity distributions of the biota could be explained by five principal components corresponding to five overlapping salinity zones: 0-4 ppt; 2-14-ppt; 11-18-ppt; 16-27-ppt; and $\geq$24-ppt. This zonation scheme is similar to the Venice system, but is objectively derived from the salinity distribution of estuarine organisms. Measurement of the ionic strength of estuarine and marine waters is typically made using salinity. Salinity may be defined as the total solids in water after all carbonates have been converted to oxides, all bromide and iodide have been replaced by chloride, and all organic matter has been oxidized (APHA 1981), and is usually reported as grams per kilogram or parts per thousand. Salinity is most commonly measured electronically using a salinometer probe as part of a CTD unit.

A related measure of the ionic strength of water samples is the conductivity, which is the ability of an aqueous solution to carry an electric current. This ability depends on the presence of ions, their total concentration, mobility, valence, relative concentrations, and on the temperature of measurement.

Conductivity is a more useful measure in the tidal fresh water portion of estuaries than is salinity (or chlorinity). Conductivity is most frequently measured using a CTD meter.

The EMAP-Estuaries program collects point-in-time salinity measurements concurrently with the collection of biological and sediment samples using a CTD probe (Holland 1990). CTD-measured salinity is also incorporated in other estuarine monitoring programs; for example, the Chesapeake Bay (Holland et al. 1988, 1989), San Francisco Bay (ABAG 1991), and Puget Sound (PSWQA 1988, 1990, 1991). Monitoring guidance for the National Estuary Program (USEPA 1992) and procedural and monitoring guidance for the CWA §403 program (USEPA 1994a) both recommend CTD probes as the preferred method for collecting salinity data.

3.3.2   Temperature

Temperature is an important determinant of the rate of chemical reactions and biological processes. DO saturation is a function of water temperature. Temperature influences the spatial and seasonal distribution of benthic infauna (Kendall 1983 cited in Dardeau et al. 1992), microbial process rates (Christian 1989), and temporal and spatial distributions of fishes (e.g., Houde and Zastrow 1991). Estuarine water temperature in temperate regions is primarily a function of the temperatures of influent streams, rivers, the ocean, and tidal stage (Reid and Wood 1976). In the sub-tropical estuaries of Florida and Texas, estuarine temperature may be more closely related to incident sunlight and air temperature. Because most estuaries are shallow, there can be considerable diurnal and seasonal temperature variation. Estuarine temperature also varies with air temperature and depth, leading to vertical temperature gradients.

In addition to the potential influence of natural temperature variations on aquatic biota and chemical reactions, anthropogenic thermal inputs can lead to significant modifications of estuarine and coastal marine biological communities. A prime example is thermal loading via discharge of cooling water from power plants and other industrial facilities. The important influence of thermal discharges is recognized in §316 of the CWA, which allows USEPA or states to impose effluent limitations on thermal loading at point sources to ensure that balanced, indigenous populations of shellfish, fish, and wildlife in and on a water body will be maintained. Temperature should be measured at each sampling site with a CTD probe at 1-m intervals from the surface to within 1-m of the bottom concomitantly with the collection of salinity and DO data. Diel temperature measurements may also be needed.

3.3.3   Dissolved Oxygen

Dissolved oxygen (DO) is a basic physiological requirement for nearly all aquatic biota and for the maintenance of balanced populations (exceptions being anaerobic systems). Most estuarine populations can tolerate short exposures to reduced DO concentrations without adverse effects. Extended exposures to DO concentrations less than 60% oxygen saturation may result in modified behavior, reduced abundance and productivity, adverse reproductive effects, and mortality (Holland et al. 1989, Reish and Barnard 1960, Vernberg 1972). Low DO conditions can also increase the vulnerability of benthos to predation as they extend above the sediment surface to obtain more oxygen. Exposure to less than 30% saturation ($\sim$2-mgL^{-1}) for 1 to 4 days causes

mortality to most species, especially during summer when metabolic rates are high. Some benthic macroinvertebrate species tolerate low DO conditions, and prolonged low DO concentrations frequently lead to changes in the composition of benthic macroinvertebrate assemblages in certain areas (Holland et al. 1989). Aquatic biota exposed to low DO may be more susceptible to adverse effects of other stressors (e.g., disease, toxic chemicals, habitat modification) (Holland 1990).

Because DO concentration throughout the water column can vary widely with tide, time of day, wind patterns, and biological activity, the EMAP-Estuaries program conducted extensive comparisons of point-in-time and continuous collections of DO data. In deeper areas of Chesapeake Bay, within EMAP's Virginian Province, bottom DO has a strong tidal signal; high tide corresponds to lower DO near the bottom. Significant serial autocorrelation of dissolved oxygen concentration persists for 6 to 8 days, indicating that consecutive measurements taken less than 8 days apart may not be independent (Holland 1990). In some EMAP Louisianian Province estuaries a strong diurnal cycle with lower DO occurs at night (Stickney 1984, Turner et al. 1987). Low DO in these shallow, often well-mixed estuaries may be highly variable both spatially and temporally (Sanford et al. 1990, Schroeder 1979, Rabalais et al. 1985). These conditions can lead to misclassification of the ecological condition of estuaries with respect to hypoxia.

Goals of the EMAP analyses comparing DO measurement approaches were to determine the best measure for representing DO exposure, to evaluate the stability of DO over the sampling

(i.e., index) period, and to determine if the characteristics of exposure to extreme low DO can be predicted from short-term continuous DO records (Holland 1990). Analyses of the data for the Virginian Province showed that three instantaneous DO profile measurements best characterized the DO status of a site. In contrast, in the Louisianian Province the minimum DO measured over a 24-hour period was most successful in characterizing both low and high frequency hypoxia sites. These differences have logistical implications for bioassessment of estuarine and coastal marine waters in that instantaneous DO measurements can be made with CTD meters equipped with a DO probe; the 24-hour minimum DO measurements require the use of a continuous recording DO meter that must be deployed and subsequently retrieved. Consequently, dissolved oxygen is an important habitat parameter, but the manager must exercise care in both sampling design and data interpretation when attributing biotic responses to potential hypoxia.

Collection of DO in Tier 1 of these procedures should include an instantaneous measurement at the same stations and times as biological samples are collected. Tier 2 should include measurement of DO in the early morning at each station as a minimum. Tier 3 should include DO collected along a depth profile from surface to within 1-m of the bottom at 1-m intervals. For more detailed characterization of DO conditions, particularly at sites which may undergo hypoxia, recording DO meters may need be deployed in Tier 3. In any case, as the EMAP experiences indicate, careful DO profiles should be established for each region surveyed before any presumptions about community response are made.

3.3.4 pH

Another important indicator of the chemical condition of estuarine and coastal marine waters is pH. In estuaries, pH will usually be controlled by the mixing of seawater solutes with those in the fresh water inflow. A major factor influencing the pH of estuarine waters is the carbon dioxide solubility, which is a function primarily of salinity and secondarily of temperature. Seawater is a very stable buffering system containing excess bases, notably boric acid and borate salts, carbonic acid and carbonate. Surface seawater pH usually ranges between pH 8.1 and 8.3. River waters usually contain a lower concentration of excess bases than seawater; this shifts the carbonate buffering system toward a higher concentration of free carbon dioxide and lower pH in the upper reaches of rivers. Because fresh water inflow to estuaries is typically much less buffered than seawater, greater variation in pH is observed in the less saline portions of estuaries than near their mouths. The range of pH values observed in the upper reaches of estuaries can be 7.5 - 9.0.

Measurement of pH in estuaries and coastal marine waters can provide an indication of possible pollutant input (e.g., releases of acids or caustic materials) or high concentrations of phytoplankton (due to photosynthesis and respiration, pH varies inversely with the free carbon dioxide concentration and directly with DO).

3.3.5 Turbidity

The major component of turbidity in estuaries is silt. The volume of silt discharged into estuaries by streams and rivers varies seasonally, with the maximum discharge occurring during the wettest months. Silt may also be resuspended from sediments within estuaries. Turbidity has two primary effects in estuaries. First, light penetration is reduced, which directly affects primary production and abundance of aquatic macrophytes in the estuary. Second, settling of the particulate matter can result in deposition zones of mud, silt, other sediments, and detritus. This deposited material can cause changes in the composition of benthic invertebrate assemblages. For example, deposition of mud and silt can result in the clogging of gills of oysters and other filter-feeding species and a loss of a hard substrate required by these species. In coastal areas, the deposition of silt in pockets of uneven sandy bottom contributes to the "patchy" distribution of benthic invertebrate species, especially annelids, amphipods, and isopods. Deposited material can also contain particle-adsorbed contaminants; this can result in contaminated sediment "hot spots". In contrast to these negative effects on the benthic invertebrate assemblage, turbidity can have positive effects on the fish assemblage by increasing protection from predators by reduced visibility.

Turbidity can be easily assessed (as light penetration) using a Secchi disk, which is probably the most widely used method for estimating light penetration (USEPA 1992). Secchi disks are easy to use, the results are easy to interpret, and they are suitable for estimating light attenuation coefficients through the water column. Secchi disk measurements may vary somewhat because of interpersonal differences in visual acuity of observers, and, therefore, caution must be exercised when comparing Secchi disk readings taken by different investigators.

If measurement of turbidity *per se* is deemed necessary by a state, it may be accomplished *in situ* by using a

transmissometer or turbidimeter as part of a CTD system. Nephelometry is the preferred method for measuring turbidity because it is more sensitive, effective over a wider range of turbidities, and less sensitive to particle size variations than other methods (USEPA 1992).

3.3.6 Nutrients

Nutrient fluxes are central in controlling the primary and secondary production of estuaries. Estuarine autotrophs (i.e., algae, diatoms, vascular plants) require numerous macro- and micronutrients and vitamins, including C, N, P, Si, S, K, Mg, Na, Ca, Fe, Mn, Zn, Cu, B, Mo, Co, V, thiamin, cyanocobalamin, and biotin (Hutchinson 1967 cited in Day et al. 1989). Higher trophic levels are influenced indirectly by nutrients through their dependence on a phytoplankton food base. Nitrogen, phosphorus, carbon, and silicon (used by diatoms) are the most studied of the nutrients in estuaries and coastal marine waters (Bricker and Stevenson 1996). This guidance document focuses particularly on nitrogen and phosphorus as the two key, potentially limiting and more manageable nutrients for the autotrophic assemblages; i.e., macrophytes, phytoplankton, to be incorporated in bioassessment procedures. The measurement of water column nutrient concentrations in Tiers 2 and 3 will aid in identifying possible sources of biological impairment. NOAA maintains a nationwide database on eutrophication and toxic algae blooms that can be cited to provide water column information.

It is a basic tenet of botany that multiple nutrients are necessary for plant growth, and that a shortage of any single nutrient will limit further growth. Thus, estuaries are sometimes referred to as "nutrient-limited." This concept is based on the finding that, under good growth conditions, algae have a relatively stable N:P atomic ratio of about 15-16:1. This ratio is frequently known as the Redfield ratio (Redfield 1958, Correll 1987, Malone et al. 1996). The Redfield ratio is an approximation and can vary depending on the stage of algal cell division, changes in light intensity or quality, or temperature (Correll 1987). Even considering these factors, measurement of nitrogen and phosphorus concentrations in estuarine and coastal marine waters provides a useful benchmark for evaluating the possible effects of increased loadings of these nutrients. Nutrient limitations in estuarine and coastal marine waters may change seasonally in response to temporal variations in nutrient loadings in the watershed and in hydrologic patterns. In the Chesapeake Bay, phytoplankton appear to be P-limited during spring when biomass reaches its annual maximum and N-limited during summer when phytoplankton growth rates are highest (Malone et al. 1996). In North Carolina estuaries, N-limitation occurs across a trophic gradient of highly productive (Pamlico), moderately productive (Neuse), and less productive systems (Beaufort, Morehead City area) (Mallin 1994). Phosphorus may be co-limiting in some of these areas during portions of the year (Mallin 1994). In the Florida Keys and adjacent Florida Bay, LaPointe and Clark (1992) determined that phosphorus is the primary limiting nutrient, with nitrogen being co-limiting. Smith and Hitchcock (1994) concluded that phosphorus (and silica) potentially limits phytoplankton growth in the Louisiana Bight during winter-spring, particularly at low salinities. In late summer nitrogen may be limiting in higher salinity waters of the Louisiana Bight.

Nitrogen and phosphorus occur in estuarine and coastal marine waters in

many forms which can be variously described in terms of oxidation state, phase (solid-liquid-gas), chemical structure, or analytical method. Nitrogen forms are the most diverse, with nitrogen compounds ranging from NO_3^- to NH_4^+. Dissolved nitrogen species that could be incorporated into chemical analyses of this nutrient include total dissolved N (TDN), and dissolved inorganic nitrogen (DIN = $NH_4^+ + NO_3^- + NO_2^-$) (LaPointe and Clark 1992). Nitrate concentrations are typically controlled largely by external inputs to estuarine and coastal marine waters via land runoff. In some areas (e.g., Chesapeake Bay) atmospheric deposition may account for an important fraction of the nitrogen load to the water body (Dickerson 1995, Boynton et al. 1995). Ammonia concentrations are highest in waters receiving large inputs of sewage (Day et al. 1989). Dissolved organic nitrogen (DON) can be calculated as TDN minus DIN (LaPointe and Clark 1992).

Measurements of total dissolved P (TDP) and soluble reactive phosphorus (SRP) can be used to estimate dissolved organic phosphorus (DOP = TDP - SRP) (LaPointe and Clark 1992).

3.3.7 Contaminants

Measuring organic compounds and metals is particularly important because of the adverse effects they can have on aquatic life and on human health and recreation if these contaminants enter the food chain. Sources of organic and inorganic chemical contaminants include direct release to the water body, urban runoff, atmospheric deposition, industrial and municipal discharges, and upstream runoff (Velinsky et al. 1994, Wade et al. 1994). Organic and metals contaminants in the water column will usually be adsorbed onto sediment particles, settle to the bottom,

and become a source of toxicity to organisms and bioaccumulation to the food chain. Contaminant analyses should be tailored to the types of substances known or suspected as chemicals of potential concern at a site. Chemical concentrations should be compared to applicable sediment quality guidance documents to aid in interpretation and to provide an effort-based assessment method. Results of any toxicity tests conducted should be compared against results from controls and against statistical standards to provide relative rankings. Chemical analyses, toxicity tests, and benthic analyses constitute the sediment quality triad, originally proposed by Long and Chapman (1985).

The Sediment Quality Triad approach, (SQT) (Long and Chapman 1985, Chapman et al. 1987, Long 1989, Chapman 1996) can be used to assess pollution-induced estuarine and coastal marine system degradation (Schlekat et al. 1994). In an analysis of sediment metals concentrations from 497 sites in Gulf of Mexico estuaries, Summers et al. (1996) normalized metals concentrations for extant concentrations of aluminum to identify the concentrations expected from natural sources versus anthropogenic sources. Krumgalz (1993) applied a "fingerprints" approach to estuarine and coastal marine pollutant source identification. This approach assumes that if anthropogenic pollutants in a particular area had originated from the same source, then pairwise relationships between the concentrations of these pollutants in sediments from various sampling sites in the contaminated area would be linear. The correlations between pollutants will depend on the origin of the contaminants and on the patterns of mixing contaminated sediments and contaminants with "pure" sediment. Thus, the "fingerprints" can be used to

trace the distribution of contaminants from a source, or conversely, to identify potential sources.

3.3.8 Depth

Depth characterization is important for evaluating DO, temperature and salinity profiles, tidal regime consistency, and the percent of the water column that is photic. This may be especially significant in coastal areas where bathymetry changes can be great and other potentially related distinctions such as grain size are not as evident.

3.4 Bottom Characteristics

The SQT approach is generally the most comprehensive assessment of relative sediment quality. In this approach data are collected to determine concentrations of potentially toxic chemicals in sediments, to measure the relative bioavailability and toxicity of sediment-associated toxicants with laboratory bioassays, and to identify degradation of resident infauna possibly attributable to the contaminants.

Chemical data can be generated through analyses of the bulk sediments and compared to applicable sediment quality guidelines (SQGs), proposed or promulgated criteria or standards wherever they exist, and relevant sedimentological factors such as grain size, total organic carbon, or aluminum. Data analyses can be conducted to identify both sites and chemicals of potential concern. Chemicals of highest potential concern are those in which SQG or other applicable values were exceeded most frequently and by the largest amount. These chemicals also would be expected to show strong associations with measures of toxicity. Moreover, chemicals of highest concern may be those determined to be bioavailable and bioaccumulative in

laboratory exposures (Chapman et al. 1988). If possible, historical data should be used to focus the chemical analyses.

Toxicity data can be generated through a battery of short-term tests performed under the controlled environment of the laboratory. Amphipod survival tests, which are most frequently used in North America, are done with exposures to solid-phase sediments and percent survival is measured after 10-days. These acute tests often are accompanied by tests of other sediment phases (e.g., pore waters and solvent extracts) with sublethal endpoints (Long et al. 1996). Data from the tests provide information for ranking and prioritizing sampling sites according to their potential for causing adverse effects among resident infauna. Confirmation of adverse effects among the infauna must be done with analyses of samples collected from the same locations tested for chemical concentrations and toxicity. Measures of species richness, total abundance, relative abundance of crustaceans (particularly infaunal amphipods) and/or other relatively sensitive taxa provide information on the degree to which resident biota have been adversely affected. Chapman (1996) provided useful guidance on the interpretation of the SQT data.

Measurements of bottom characteristics of estuaries and coastal marine waters provide important data for interpreting the condition of targeted biological assemblages. Sediment grain size influences the spatial distribution of benthic macroinvertebrates and fishes. Fine-grained sediments can adsorb contaminants, creating a source of potential impairment to bottom communities. Organic carbon found in these sediments can mediate the concentrations of DO, organic contaminants, and metals. Measuring the depth of the sediment oxidation-

reduction potential discontinuity layer provides information on aerobic vs. anaerobic respiration in sediments. Sediment nutrients such as particulate nitrogen and phosphorus can be re-mobilized by physical disturbance or changes in the water column chemistry to become an additional nutrient source leading to potential eutrophication, phytoplankton blooms, and hypoxia of estuarine and coastal marine waters. Finally, sediment contaminant measurements can provide insights on factors that might limit biological assemblages and lead to potential human health effects.

3.4.1 Sediment Grain Size

The objective of measuring sediment grain size is to detect and describe spatial and temporal changes of the benthic habitat. The availability of sediment contaminants is often correlated with sediment grain size because more sediment contaminants are adsorbed onto small grained sediments due to their greater surface area. Likewise, grain size information may explain the temporal and spatial variability in biological assemblages related to an organism's ability to build tubes, capture food, and escape predation. Grain size data may be used to determine the extent of or recovery from environmental perturbations, to evaluate the condition of benthic habitats, and to assist in providing early warnings of potential impacts to the estuarine ecosystem (USEPA 1992). The most common measurements and classifications of sediment grain size are as follows:

- ‣ clay   < 0.004-mm
- ‣ silt   0.004 - 0.064-mm
- ‣ sand   0.064 - 1.0-mm
- ‣ gravel   > 1.0-mm

3.4.2 Total Organic Carbon, Total Volatile Solids, and Acid Volatile Sulfides

Total organic carbon (TOC) and acid volatile sulfides (AVS) are considered by some to be the most important sediment properties determining the bioavailability and toxicity of certain organic compounds and trace metals in sediments (DiToro et al. 1990, DiToro et al. 1991). The importance of these factors is based on an equilibrium partitioning approach. This approach assumes that the bioavailable fraction of chemicals in sediment is correlated to that fraction in the porewater rather than whole sediment concentrations. Therefore, factors that influence the partitioning of compounds between sediment and porewater will govern bioavailability. In addition, it is assumed that equilibrium exists among the phases (hence, the name "equilibrium partitioning"). For non-ionic hydrophobic organic chemicals, the primary factor influencing partitioning is TOC; for certain divalent cationic metals; i.e., cadmium, copper, nickel, lead, zinc, an important binding phase is the acid volatile sulfide fraction. The development of sediment quality criteria by USEPA is based on these assumptions and a comparison of predicted porewater concentrations to existing water quality criteria (DiToro et al. 1991, Ankley et al. 1996).

Normalization of non-ionic organic compounds is accomplished by calculating chemical concentrations per gram of sediment organic carbon rather than per gram of dry sediment. This approach allows comparisons of the potential bioavailability of non-ionic organic compounds across different sediment types and can be used to screen for chemicals of concern. For an explanation for how to apply this approach to calculate sediment quality

Habitat Characterization

criteria the reader is referred to DiToro et al. (1991).

Although SQGs (Sediment Quality Guidelines) derived for organic compounds and trace metals reported on a dry weight (bulk) basis have been shown to be reliable and predictive of both non-toxic and toxic conditions (Ingersoll et al. 1996, MacDonald et al. 1996, Long et al. 1998a); these values do not account for the relative bioavailability of sediment-associated chemicals. In areas in which high trace metal concentrations are known or suspected, or in which SQGs are exceeded, further evaluations of chemical contamination may be aided in subsequent analyses by use of the simultaneously extracted metals/acid volatile sulfides (SEM/AVS) tool to provide estimates of bioavailability.

The AVS normalization approach assumes that select trace metals bind to sediment sulfide, specifically the sulfide fraction soluble in cold acid, known as AVS (Allen et al. 1993). The bioavailability of trace metals capable of forming insoluble metal sulfides will be determined by the proportion of these metal ions not bound to sulfide. Hence, on a molar basis, if the concentration of SEM is less than the molar concentration of AVS, all of the metals should precipitate as metal sulfides and not be bioavailable. Conversely, if SEM exceeds AVS then free metal ions may exist in the porewater. This approach appears to work best in situations when the ratio of [SEM]/[AVS] is less than 1.0 or the difference between SEM and AVS concentrations is less than 0.0 (Hansen et al. 1996). That is, the SEV/AVS tool is primarily intended for use as a no-effects tool and caution is advised in using it as a predictor of toxicity or other effects. Long et al. (1998b) reported that the SEM/AVS tool and SQGs based upon bulk sediment chemistry for trace

metals performed equally well in correctly predicting samples as either toxic or non-toxic. For an excellent discussion of the applicability, advantages and disadvantages of this approach, the reader is referred to several review papers in Environmental Toxicology and Chemistry, Volume 15, #12.

3.4.3 Sediment Oxidation-Reduction Potential

There are four oxidation-reduction (redox) processes related to biological respiration that occur in benthic sediments. These chemical reactions, which depend on the availability of electron acceptors; i.e., oxygen, nitrate, sulfate, and carbonate, stratify the sediments into four zones. Oxygen is the electron acceptor used for aerobic respiration and is the most important oxidizing agent at the surface of the sediments. Nitrate reduction occurs between 0- and 4-cm and produces elemental nitrogen. This is followed by sulfate reduction which produces hydrogen sulfide. Carbonate reduction occurs between 10- and 50-cm and results in the production of methane.

Aerobic respiration will be the dominant reaction as long as oxygen is available. The depth at which oxygen is fully depleted and the redox potential goes to zero has been termed the redox potential discontinuity (RPD) layer (Day et al. 1989). Burrowing organisms oxidize the sediments and hence will increase the zone of habitability as reflected in the depth of the RPD layer. Color changes in the substrate occur as a result of the oxidation and reduction of metals, such as iron, in the sediments. The upper few centimeters may appear brown from the formation of iron oxides and hydroxides, whereas the zone of reduction turns gray and eventually

black in the deeper sediments from the formation of ferrous sulfide and pyrite. When examining a cross-section of a sediment core sample, the RPD layer is visibly noticeable by this change in color. The depth of this color change should be recorded because, as noted above, it indicates the zone of habitability for benthic infauna. The closer to the sediment surface this color change appears, the less available dissolved oxygen exists in the sediment porewater.

3.4.4   Sediment Contamination

Sampling the surface sediments for the presence of contaminants can provide insight on factors limiting the benthic community, as well as the potential for impacts to human health; i.e., by biomagnification or bioaccumulation in the food chain or by the contamination of shellfish. Metals and organic chemicals entering estuaries from fresh water inflows, point sources of pollution, and various nonpoint sources, including atmospheric deposition, generally are retained within estuaries and accumulate within the sediments (Forstner and Wittman 1981, Hinga 1988, Nixon et al. 1986, Schubel and Carter 1984, Turekian 1977) because of the affinity of most contaminants for particle adsorption (Hinga 1988; Honeyman and Santschi 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 upper tidal fresh water, low current velocity, deep basins, and the zone of maximum turbidity in the upper reaches of estuaries within which suspended sediment concentrations are greater than those either farther upstream or farther seaward (Schubel and Carter 1984). The concentration of contaminants in sediments is dependent upon interactions between natural (e.g., chemical and physical sediment characteristics) and anthropogenic factors (e.g., type and volume of contaminant loadings) (Sharpe et al. 1984).

Bottom sediments in some estuaries (e.g., harbors near urban areas and industrial centers) are so contaminated that they represent a threat to both human and ecological health (NRC 1989, OTA 1987, Weaver 1984), but contaminated sediments are not limited to these areas. Pollutant runoff from agricultural areas also is an important source of contaminant input to estuaries (Boynton et al. 1988, Pait et al. 1989).

The EMAP program uses the NOAA National Status and Trends (NS&T) suite of contaminants as the basis for measurements in homogenized subsamples of collected sediments (Figure 3-1). A useful citation for the NS&T Program list of chemicals is O'Connor et al. 1994. The NOAA suite includes chlorinated pesticides, polychlorinated biphenyls (PCBs), polyaromatic hydrocarbons (PAHs), major elements, and trace metals.

3.5   Proposed Habitat Parameters

Table 3-1 summarizes the proposed habitat measurements by survey tier and provides possible sources of information, methods, and equipment, as appropriate. Agency-specific objectives will determine the overall design of any sampling program. The following tier distribution is just one approach possible for gathering and organizing data. Habitat measurements are intended to be cumulative across tiers; that is, the desktop screening of Tier 0 should be incorporated into Tier 1, Tiers 0 and 1 parameters should be incorporated into Tier 2, and Tiers 0, 1,

Figure 3-1

Polyaromatic Hydrocarbons (PAHs)

Acenaphthene
Anthracene
Benz(a)anthracene
Benzo(a)pyrene
Benzo(e)pyrene
Biphenyl
Chrysene
Dibenz(amh)anthracene
2,6-dimenhyulnaphthalene
Fluoranthene
Fluorene
2-methylnaphthalene
1-methylnaphthalene
1-methylphenanthrene
Naphthalene
Perylene
Phenanthrene
Pyrene
Benzo(b)fluoranthene
Acenaphthlyene
Benzo(k)fluoranthene
Benzo(g,h,i)perylene

Major Elements
Aluminum
Iron
Manganese
Silicon

Trace Elements
Antimony
Arsenic
Cadmium
Chromium
Copper
Lead
Mercury
Nickel
Selenium
Silver
Tin
Zinc

DDT and its metabolites

o,p'-DDD
p,p'-DDD
o,p'-DDE
p,p'-DDE
o,p'-DDT
p,p'-DDT

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

18 PCB Congener

Congener Number	Location of Cls							
8	2	4'						
18	2	2'	5					
28	2	4	4'					
52	2	2'	5	5'				
44	2	2'	3	5'				
66	2	3'	4	4'				
74	2	4	4'	5				
99	2	2'	4	4'	5			
101	2	2'	4	5	5'			
118	2	3'	4	4'	5			
153	2	2'	4	4'	5	5'		
105	2	3	3'	4	4'			
138	2	2'	3	4	4'	5'		
187	2	2'	3	4'	5	5'	6	
128	2	2'	3	3'	4	4'		
180	2	2'	3	4	4'	5	5'	
170	2	2'	3	3'	4	4'	5	
195	2	2'	3	3'	4	4'	5	6
206	2	2'	3	3'	4	4'	5	5'	6
209	2	2'	3	3'	4	4'	5	5'	6	6'

Other measurements
Tributyltin
Acid volatile sulfides
Total organic carbon

Figure 3-1

Chemicals measured in sediments by the EMAP-Estuaries program. Refer also to the NS&T Program list of chemicals (O'Connor et al. 1994).

and 2 parameters should be incorporated into Tier 3. The habitat tiers described here should be used with the corresponding biological survey tier.

3.5.1 Tier 0

The purpose of a Tier 0 assessment is to support the planning for monitoring and more detailed assessments. Tier 0 is a desktop screening assessment in which documented information for the estuary or coastal marine areas of concern is compiled from sources including databases, peer-reviewed and gray literature, state and federal agencies, universities, and local experts. A Tier 0 assessment should always precede any of the three subsequent tiers. Examination of long-term data records (e.g., salinity, DO, climate) is particularly important for identifying the variability which must be accounted for in the design of subsequent field monitoring. Habitat parameters to examine in Tier 0 include:

Area and Geomorphometric Classification

The size and classification of an estuary indicates its potential to respond to various impacts. Estuary types include coastal plain, lagoon, fjord, and tectonically-caused. Circulation type (e.g., gravitational, tidal, wind-induced) influences current patterns, salinity regimes, and thermal and dissolved oxygen patterns.

Habitat Type

Identifying and delineating the various habitat types (Section 3.3) that occur in the estuary or coastal marine waters will be necessary for partitioning the natural variability in the system. The extent of such a delineation will depend on the size of the area of concern and the nature of the environmental gradients

present. Initial partitioning will probably be based on salinity, sediment type, and depth.

Watershed Land Use and Population

Land use and population data for the watershed will help to identify classes of contaminants and other stresses that may affect the water body. For example, agricultural areas located near an estuary might be expected to be sources of nonpoint loading of nutrients, pesticides, herbicides, and sediment. Urban areas may contribute toxic compounds via stormwater runoff. The pattern and magnitude of population density in the watershed can potentially provide clues regarding the potential for human-induced impacts to the water body.

Water Column and Bottom Characteristics

Historic data on water column and bottom characteristics is central for identifying system variability and to support the design of subsequent monitoring. This data can also be used by states to identify types and locations of potential impairment, for example, areas with high concentrations of water column nutrients, suspended sediment, or sediment contaminants.

3.5.2 Tier 1

Tier 1 is a basic field assessment that is used for screening purposes to identify potential reference and impaired sites. For biocriteria development purposes, it is adequate for only rudimentary habitat classifications and evaluations. It identifies the general physical characteristics of the estuary or coastal region, the habitats, and the potential sources of anthropogenic stress. Tier 1 relies heavily on existing information identified in Tier 0 and supplemented

by one-time easily-measured field parameters, which are measured concomitantly with the collection of biological data. Much of the habitat information needed in this tier can be acquired from state or federal agency records. Depending on the needs of the state, Tier 1 habitat characterizations may include the following elements:

Estuary Characteristics

Information on estuary characteristics is essential for the development of appropriate sampling strategies. Data to be compiled in this category includes estuary area, geomorphometric classification (e.g., classical coastal plain estuary, lagoon), and habitat types present (e.g., hard bottom, soft bottom). These data can be obtained from USGS maps, NOAA charts, and reports and data archives at federal and state agencies and universities.

Watershed Characteristics

Knowledge of watershed characteristics can provide important information for determining appropriate sampling station locations and for evaluating possible sources and causes of biological and habitat impairment. Data to be compiled for this category include watershed land use, human population density, NPDES discharge locations, and vegetative cover. These data are likely to exist at federal, state, and local agencies and universities.

Water Column Characteristics

The characteristics in a water column play a key role in determining the biota present at a given location and their broader distribution patterns. Parameters to be measured in Tier 1 in the field include salinity and conductivity, DO, temperature, pH, turbidity and Secchi depth. Records of

tidal stage and current velocity at the time of sampling at each station should be acquired from NOAA.

Bottom Characteristics

Characteristics of bottom sediments also are key determinants of aquatic biota present in estuaries and coastal marine waters. Parameters to be measured in Tier 1 include depth, dominant sediment type, total volatile solids, and sediment RPD layer depth.

3.5.3 Tier 2

Tier 2 provides the information necessary to develop a quantitative ranking of the sites that can be used to prioritize resources and sampling efforts. Habitat information collected in Tier 2 will be used in the development of biocriteria. In addition to Tier 1 habitat characterization, Tier 2 includes:

Water Column Nutrients

Nutrients in the water column determine the nutrient state of the area and may indicate possible sources of impairment, particularly nonpoint source runoff.

Sediment Characterization

The organisms closely associated with the bottom are strongly influenced by such sediment characteristics as the average grain size and the percent composition of silt, sand, and clay (Day et al. 1989). These characteristics determine the structure of the benthic community based on the preferences of the major groups of organisms. For example, suspension feeders are found more often in firmer, sandier substrates than are deposit feeders; interstitial meiofauna are predominant in sandy areas, whereas burrowing meiofauna prefer silty mud. Although a high

organic content in the sediments can increase the rate of oxygen depletion, there are many organisms that require high concentrations of organic matter. Sediment characteristics to be measured in Tier 2 include percent sand vs. silt-clay, mean grain size, total volatile solids, and total organic carbon.

Shorezone Vegetative Cover Characterization

Shorezone vegetation provides stability for beaches, wetlands, banks, and cliffs, serving to reduce erosion and nonpoint source runoff to the water body. As such, evaluation of shorezone vegetative cover is important for identifying possible sources of impairment and remedial approaches. Terrestrial riparian vegetated areas to consider are uplands and the floodplain. Areas of emergent, intertidal, and submerged vegetation should also be characterized.

Shorezone vegetative cover is important for reducing nutrient and sediment loading to estuaries from nonpoint source runoff, attenuating incident wave energy and reducing shore erosion, and providing important nursery and feeding habitat for migratory species. Salt marshes and aquatic macrophytes have high gross primary productivity and provide a source of autochthonous organic matter for detrital feeders in adjacent waters. An assessment of the coverage and types of shorezone vegetation can contribute to the overall assessment of the condition of estuarine and coastal marine habitat. Evaluation of vegetative cover is most easily accomplished by aerial photography and mapping coupled with ground-truthing. Detailed procedures used for photography and mapping aquatic macrophytes are provided by Orth et al. (1993) (Chesapeake Bay), Ferguson and Wood (1994) (North Carolina estuaries), and USEPA (1992) (National Estuary Program) and can be adapted for use in other areas.

3.5.4 Tier 3

Tier 3 provides a detailed assessment with a high level of certainty of the biological or habitat condition of the estuarine and coastal marine environment. It is the definitive assessment level to distinguish habitat variation from anthropogenic impacts when the biocriteria have been exceeded. Tier 3 focuses on biological community level investigations and thoroughly integrates the physical, chemical and biological data to yield a detailed impact assessment. In addition to the habitat parameters compiled in Tier 0 and measured in Tiers 1 and 2, sediment oxidation-reduction potential, sand/silt/clay proportions, sediment contaminants, and water column pesticides, herbicides, and metals, nutrient speciation, and AVS/SEM as needed may be measured in this tier.

Tier 3 provides the detailed diagnostic information necessary for: (1) identifying specific problem sources in the drainage area; (2) delineating mitigation options for the identified problems; and (3) preparing written management plans for the estuary or coastal marine area of interest. Although the data collected in Tier 3 cannot prove cause-and-effect relationships between identified stressors and ecosystem responses, they can provide a strong correlation and a definitive assessment, with a high degree of certainty, of the biological integrity of the target waters and their habitats.

Table 3-1. Habitat measurements for estuaries and coastal marine waters.

Habitat Measurements	Assessment Tier 0	1	2	3	Information Source	Method(s) and Equipment
Historical Information						
Estuary area	✓	✓	✓	✓	USGS quad maps	Can be estimated from maps or using GIS
Geomorphometric classification (classical coastal plain estuary; salt marsh estuary; lagoon; fjord; tectonically-caused estuary)	✓	✓	✓	✓	USGS quad maps	
Habitat type (seagrass beds; hard bottom; soft bottom; water column; emergent marsh; mudflat; sandflat; gravel/cobble; rocky)	✓	✓	✓	✓	NOAA bathymetry charts; historic surveys by federal, state agencies and universities	
Watershed land use	✓	✓	✓	✓	USGS land use maps; state planning agencies; local zoning agencies	Can be estimated from maps or using GIS
Population density	✓	✓	✓	✓	US census data	
NPDES discharges	✓	✓	✓	✓	State water quality agency	
Vegetative cover	✓	✓	✓	✓	Historic surveys by federal, state agencies and universities	Can be estimated from maps or using GIS
Field-collected Information: Water Column Characteristics						
Salinity, conductivity	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Conductivity cells (CTD meters)
DO	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	CTD meter equipped with O_2 probe, recording DO meters for Tier 3 may be needed
Temperature	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	CTD meter, satellite remote sensing.
pH	✓*	✓	✓	✓	Historic data from federal, state agencies and universities.	CTD meter equipped with pH probe

Table 3-1 (Cont'd).   Habitat measurements for estuaries and coastal marine waters.

Habitat Measurements	Assessment Tier 0	Assessment Tier 1	Assessment Tier 2	Assessment Tier 3	Information Source	Method(s) and Equipment
Turbidity, Secchi depth	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Secchi disk; transmissometer, nephelometer, turbidimeter if desired
Nutrients (nitrogen species, phosphorus)	✓*		✓	✓	Historic data from federal, state agencies and universities	Spectrophotometry
Organics, metals	✓*			✓	Historic data from federal, state agencies and universities	Standard methods for selected suite of contaminants
Field-collected Information: Bottom Characteristics						
Depth	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Meter wheel, fathometer, hydrostatic pressure sensor
Sediment grain size	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Sieving or separatory column combined with pipette analysis for silt-clay fraction
Total volatile solids	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Standard methods
Total organic carbon	✓*		✓	✓	Historic data from federal, state agencies and universities	Standard methods
Acid volatile sulfides	✓*			✓	Historic data from federal, state agencies and universities	Standard methods
Sediment reduction-oxidation potential	✓*	✓	✓	✓	Historic data from federal, state agencies and universities	Visual determination of the depth of change in sediment color in core
Sediment contamination	✓*			✓	Historic data from federal, state agencies and universities	Standard methods for contaminants selected

*   Historic data should be included in Tier 0; the tier does not include any field collection of new data.

Chapter 4

Physical Classification and the Biological Reference Condition

Estuaries and coastal marine waters span a range of spatial scales from small subestuaries, embayments, and coastal lagoons to large estuaries (e.g., Chesapeake Bay, Puget Sound) and open coastal waters. The procedures described in this document rely on a spatial hierarchy to accommodate the potentially large range of water bodies that states may assess. The top level in the hierarchy is a geographic region containing comparable landform and climate. The provinces used by the EMAP-Estuaries program (e.g., Carolinian, Columbian) are examples of this hierarchical level. The next level consists of individual watershed characteristics. Key attributes to consider at this level include land cover, the watershed-to-basin area ratio, and the geology and soils of the watershed. Examples of the use of this hierarchical level in estuarine assessment are the Chesapeake Bay watershed, New Jersey coastal bays, or California saline lagoons. The lowest level in the hierarchy considers habitat characteristics. As discussed in Chapter 3, the three primary variables used to partition spatial heterogeneity at this level are sediment grain size, salinity, and water depth. Description of sampling sites as "low mesohaline, mud" or "10-m depth, gravel" would be examples of this level of the hierarchy.

Reference conditions are expectations of the status of biological communities in the absence of anthropogenic disturbances and pollution, and are usually based on the status of multiple reference sites. Ideally, reference sites are minimally impaired by human pollution and disturbance. The care that states use in selecting reference sites and developing reference condition parameters, together with their use of standardized survey techniques, will directly influence the quality of the resulting water body assessment. At a minimum, reference conditions should be identified for each of the estuary and coastal marine classification categories developed by a state.

Reference conditions reflect the biotic potential for estuaries and coastal marine waters if they are not impaired by human activity or pollution. Attainment of an aquatic life designated use is evaluated against the reference condition as a key element in the biocriteria for that aquatic life use. Biocriteria may be set higher than the best conditions observed in the data available for an area that is highly impaired. In this instance, interim, incremental criteria may be established as the regional authority works on environmental recovery.

4.1   Classification Approach

The biological reference condition must be determined separately for each estuarine or coastal marine physical class. Assessing biological condition requires reference conditions for comparison and for development of models and indexes to help establish biocriteria and detect impairment. There is no single "best" classification nor are resources available to determine all possible differences between all

estuarine and coastal marine sites in a region. The key to classification is practicality within the region or state in which it will be applied; i.e., local conditions determine the classes. Classification will depend on regional experts familiar with the range of estuarine conditions in a region as well as the biological similarities and differences among the assessment units. Ultimately, physical classification may be used to develop a predictive model of those estuarine and coastal marine characteristics that affect the values of the biological metrics and indexes at reference sites.

The regional differences in estuarine and coastal marine biological communities across the United States must be accounted for in the development of a biological criteria program. These differences can be identified by comparing the biology of water bodies of interest to a reference condition. As biological conditions change across the country, the reference conditions will also change. To account for the regional geographic differences that create structural differences in biological habitat (either natural or human-induced), states should classify estuaries and coastal marine waters or segments thereof into groups. A reference condition should be established for each of these classification groups. Biotic index comparisons can then be made within each classification group and inappropriate biological comparisons between different classes will be precluded. Moreover, the aquatic life expectations of water bodies are tempered by realistic expectations. With biological systems, it is not possible to set uniform, nationwide numeric biological criteria.

Estuaries vary widely in size, shape, and ecological and physical characteristics, and a single reference condition that applies to all estuaries (or coastal marine waters) would be inappropriate. The purpose of classification is to group similar estuarine or coastal marine sites together; i.e., to prevent the comparison of apples and oranges. Classifying the variability of biological measures within groups inevitably requires professional judgment to arrive at a workable system that separates clearly different systems, does not consider each estuary or subestuary a special case, and does not lead to the proliferation of classification groups. The intent of classification is to identify the smallest number of groups of estuarine or coastal marine categories that under ideal conditions would have comparable biological communities for that region. As much as possible, classification should be restricted to those characteristics of estuaries and coastal marine waters that are intrinsic, natural, reasonably stable over time, and not the result of human activities.

The approach to reference condition characterization and classification is illustrated in Figure 4-1. An idealized biological potential for estuarine sites is expressed, for instance, by a fish index and an infaunal index, each within a certain range of values (Figure 4-1). A test site is compared to the expected ranges of values, and if its indexes are outside those ranges, it is judged as not meeting expectations to some degree. Test sites are usually not compared to a theoretical ideal, but to biological criteria derived from a population of reference sites. Test sites are judged as not meeting the criterion if they are beyond some predetermined limit of the distribution of reference values.
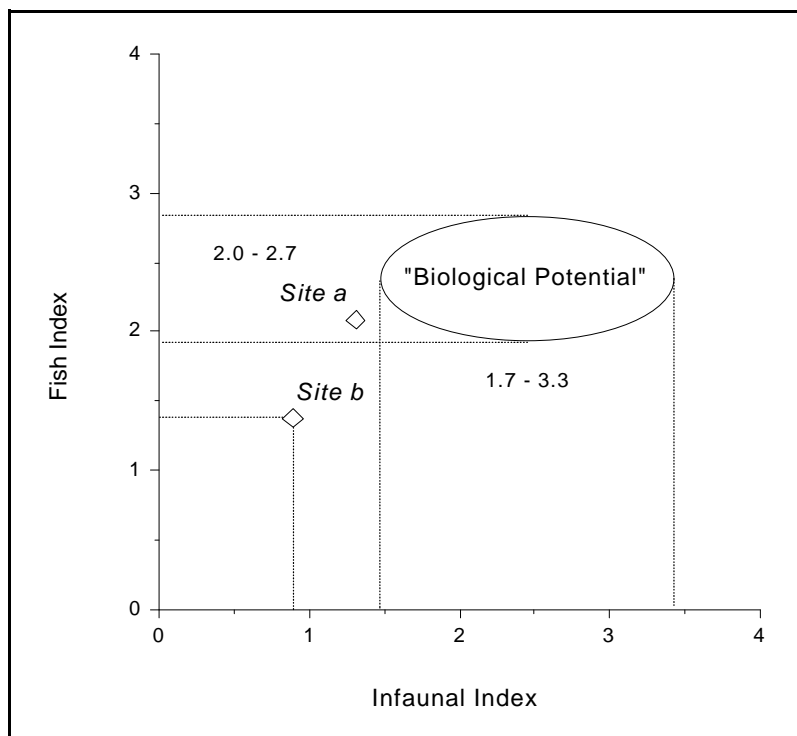
Figure 4-1

Graphical representation of bioassessment. Assessment sites *a* and *b* are compared to an ideal biological potential. Site *a* is near its potential. Site *b* deviates from it.

A population of reference sites might consist of sites which overlap different classes of estuarine or coastal marine waters (Figure 4-2). A useful classification system in this instance separates these reference sites into classes with different biological expectations. The classification itself must be based on abiotic information that is minimally affected by human activities (e.g., ecoregion, estuary and coastal marine physical characteristics, basin characteristics), such that test sites can be assigned to one of the classes before any biological information is obtained. Furthermore, the classification must explain biological variability in the reference sites (Figure 4-2). Separation into classes then lowers inherent variation and allows greater precision in assessing test sites. If test site "a" in Figure 4-2 is a member of class II, it would be judged as not meeting reference expectations. If, however, the physical classification were not done, site "a" would be judged to meet reference expectations because it is within the limits of all reference sites.

Sequence of Classification and Characterization

The general sequence of reference condition characterization is to first make a preliminary physical classification of estuaries and coastal marine areas within a region (Conquest et al. 1994). Because of natural variation among and within estuaries and coastal marine waters, reference conditions will likely differ with geographic regions, major salinity zones, depth profiles, and bottom sediment types. Following classification, reference conditions are characterized using some combination of reference sites, historical data, expert opinion, and empirical models. A key element is the use of reference sites because they represent realistic, achievable goals and can be regularly monitored. Historical data and well-documented expert opinion should be used to evaluate the information developed from the reference site data and possibly from empirical models. The preliminary classification is reconciled with the biological data to

Figure 4-2

Classification and assessment. If reference sites are not classified, Site *a* is at or near its potential. If reference sites are classified and Site *a* is in Class II, it does not meet its potential and might be judged impaired.
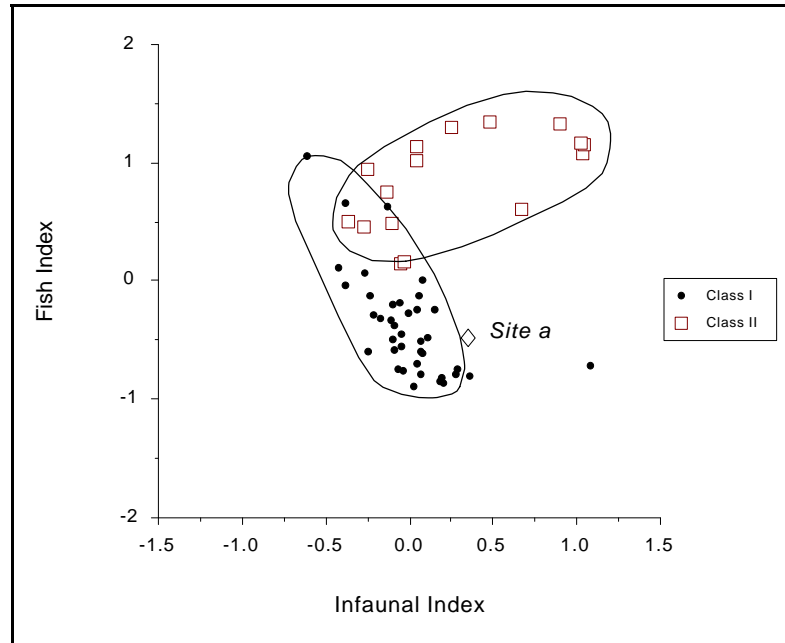
ensure that the final classification is meaningful and the reference conditions are properly characterized. The remaining sections of this chapter cover physical classification, elements of reference condition characterization, and use of reference sites. The reference site database should be periodically reviewed as data accumulate to ensure consistency of the reference characterization and classification scheme.

4.2   Physical Classification

This protocol is not intended to develop a classification scheme applicable to the entire United States. Classification within the broad estuarine categories described in Section 3.1 must be regional, and regional expertise must be used to determine those classification variables which are useful in each region.

A useful classification scheme is hierarchical, beginning at the highest (regional) level and stratifying only as far down as necessary (Conquest et al. 1994). The procedure is to classify estuaries and coastal marine waters by

geographic regions and then to increase the stratification in the classification hierarchy to a reasonable point for each given region. Although several possible classification levels are outlined below, in practice, one to three relevant levels would be entirely sufficient. Classification should avoid a proliferation of classes that do not contribute to assessment. The proposed hierarchical scheme below applies to both estuarine and coastal marine waters.

4.2.1   Geographic Region

The geographic region, be it ecoregion, physiographic province or other delineation, determines landscape-level features for classification such as: climate, topography, regional geology and soils, biogeography, and broad land use patterns. Ecoregions are based on geology, soils, geomorphology, dominant land uses, and natural vegetation (Hughes and Larsen 1988, Omernik 1987) and have been shown to account for the variability of water quality and aquatic biota in several freshwater areas of the United States. Seventy-six ecoregions were originally

identified in the conterminous United States (Omernik 1987); but recent refinements have yielded a greater resolution for some areas.

It should be noted that many of the characteristics that can be used as classification variables are often subsumed by the geographic region. For example, watersheds are often similar within major geographic regions, having resulted from the regional geomorphology. Within such regions, it might be sufficient to classify using only morphology such as depth, area, or bathymetry. Examples are the coastal bays of the Delmarva peninsula or the sounds behind North Carolina's Outer Banks.

The EMAP-Estuaries program uses biogeograpical provinces, defined by: major climatic zones and prevailing ocean currents. EMAP coastal areas in the continental United States are encompassed within seven provinces described as Acadian, Virginian, Carolinian; West Indian; Louisianian; Californian; and Columbian (Figure 4-3) (Holland 1990). These roughly approximate the traditional descriptors of New England, Mid-Atlantic Bight, Southeast Coastal, Caribbean, Gulf Coastal, Southwest, and Northwest Pacific Coast. For strictly coastal waters, this may be a sufficient level of classification.

4.2.2 Estuarine Categories

Estuaries can be categorized into four major classes based on their geomorphology: (1) coastal plain estuaries (Chesapeake Bay; Cape Canaveral, FL), (2) lagoons (Pamlico Sound, NC), (3) fjords (Puget Sound), and (4) tectonically-caused estuaries (San Francisco Bay) (Day et al. 1989). While these classifications appear to be large scale in nature, they can be used to make initial divisions of estuaries on a regional scale.

There are two types of coastal plain estuaries: classical and salt marsh. The classical coastal plain estuary is sometimes referred to as a "drowned river valley." These estuaries were formed during the last eustatic rise in sea level and they exhibit geomorphological features similar to river channels and floodplains. The salt marsh estuary lacks a major river source and is characterized by a well-defined tidal drainage network, dendritically intersecting the extensive coastal salt marshes (Day et al. 1989). Exchange with the ocean occurs through narrow tidal inlets which are in a constant state of flux. Consequently, salt marsh estuarine circulation is dominated by fresh water inflow and the tides.

Lagoons are characterized by narrow tidal inlets and uniformly shallow; i.e., less than 2-m deep, open water areas. The inlets are created by the erosion of the narrow Pleistocene ridge that formed along the coast some 80,000 years ago during the interglacial stage (Day et al. 1989). Lagoons are primarily wind-dominated and they have a subaqueous drainage channel network that is not as well-drained as the salt marsh estuary.

Classical fjords, formed during the last ice age, are river valleys that were carved out by the leading ice edge of advancing continental glaciers. When the glacier receded, large rock deposits were left behind where the leading edge had stopped. Others are also a result of glacial scouring of the coast; however, these estuaries were formed in regions with less spectacular continental relief and more extensive continental shelves, therefore they are much shallower than typical fjords.

Figure 4-3

Biogeographical
provinces adapted
from Holland
(1990). A form of
preliminary
regionalization
used by EMAP-
Estuaries.
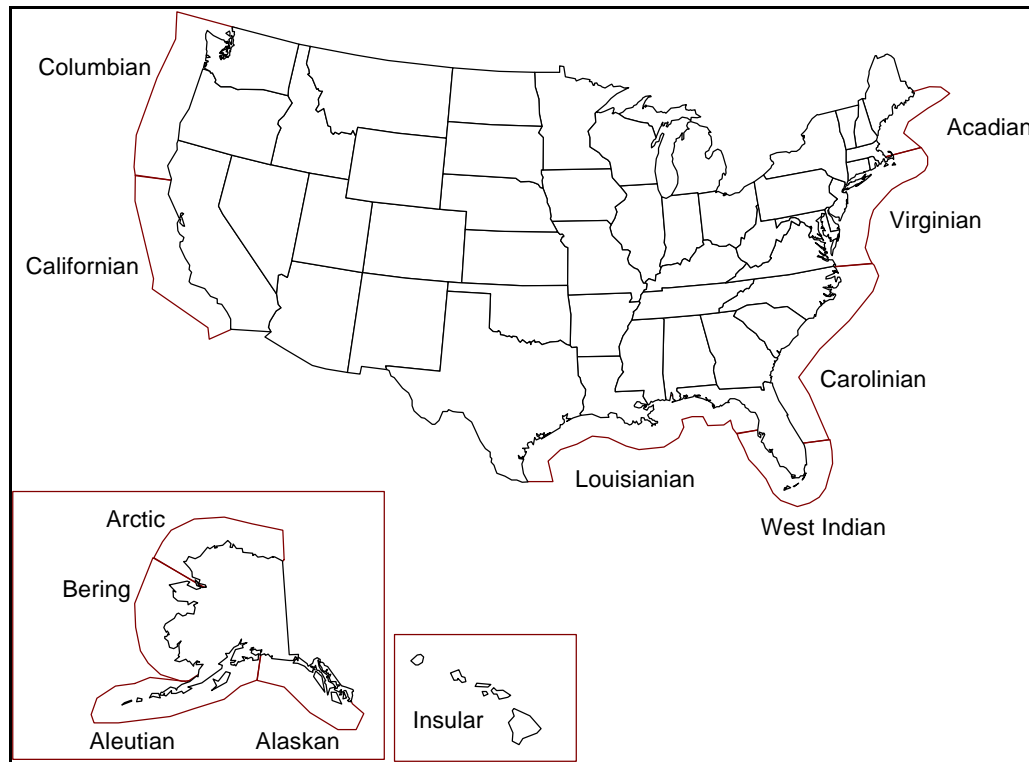
Tectonically-caused estuaries are created by faulting, graben formation, landslide, or volcanic eruption. They are highly variable and they may resemble coastal plain estuaries, lagoons, or fjords.

4.2.3   Watershed Characteristics

Watershed characteristics affect estuary and coastal marine hydrodynamics, sediment and nutrient loads, chemical and metals contaminant loads, and dissolved solids. Watershed characteristics that may be used as classification variables include:

▸   Land cover - extent of natural vegetation;

▸   Watershed-to-estuary area ratio;

▸   Soils, geology (erosiveness of soils), and topography.

4.2.4   Waterbody Characteristics

The third level of the classification hierarchy focuses on waterbody characteristics. Attributes that are considered at this level include waterbody morphology, hydrodynamics, and water quality. Each of these factors has a direct influence on the biota present in the waterbody.

Morphological Characteristics

Morphological characteristics of the estuary or coastal marine waters influence hydrodynamics and system responses to pollution. Morphological characteristics include:

▸   Depth (mean, maximum);

▸   Bathymetry - three-dimensional bottom profile;

▸   Surface area;

▸ Bottom type and sediments - substrate and grain size.

Hydrodynamics

Hydrodynamics forms a basis for water quality. Mixing and circulation patterns influence nutrient retention and the development of hypoxia.

Hydrodynamic factors include:

▸ Retention time;

▸ Stratification and mixing;

▸ Currents - speed and direction;

▸ Tidal range;

▸ Altered inflow to the waterbody, such as increased or decreased freshwater inflow from runoff or diversions.

Water Quality

As noted above, many water quality characteristics are relatively uniform within a region because they are the result of common regional, watershed, and hydrodynamic characteristics. Although water quality variables might be redundant for a classification scheme if regions are the primary classification variable, it is frequently convenient to subclassify according to water quality. An example is the practice of sub-dividing estuaries along their gradient into oligohaline, mesohaline, and polyhaline regions (see Figure 4-4 for an example of such a delineation). Water quality variables useful for classification are:

▸ Salinity and conductivity;

▸ Turbidity (Secchi depth);

▸ Dissolved oxygen (DO);

▸ pH.

Human actions (e.g., discharges, land use, freshwater flow diversions) alter water quality, especially sediment and nutrient concentrations, but they can also affect salinity, conductivity, turbidity, DO, and pH. Therefore, care must be taken that classification according to characteristic water quality reflects natural conditions and not anthropogenic impacts. For example, if estuarine sites are highly turbid due to poor land management practices in the watershed, they should not be classified as highly turbid. Instead, they should be classified according to the turbidity class they would have had in the absence of poor land use.

4.3 Establishing Biological Reference Conditions

Estuarine and coastal marine reference conditions should be established using some combination of four elements: (1) evaluation of historical data; (2) sampling of reference sites; (3) prediction of expected conditions using models; and (4) expert consensus. Each element has its inherent strengths and weaknesses (Table 4-1) that states must consider relative to their program needs, available data, and staff expertise.

4.3.1 Historical Data

In many cases, historical data are available that describe past biological conditions in the region. For the purpose of this document, historical data are datasets collected by programs that are no longer active; in many cases using methods now superseded by other methods. Careful evaluation of these data provides insight about past and potential community composition of estuarine and coastal marine waters and is an important initial phase in the biocriteria development process.

Figure 4-4

Estuarine and coastal marine biocriteria survey method useful for stratified random (population distribution) reference site selection. Wet season/high flow salinity pattern showing mainstem sampling sites for four salinity and three substrate classifications.

Wet Season/High Flow

Oligohaline

Mesohaline

Polyhaline

Marine

Key
- [ ] mud
- [▧] transitional
- [▤] sand
- [▨] coastal land area
- [●] single comprehensive survey site
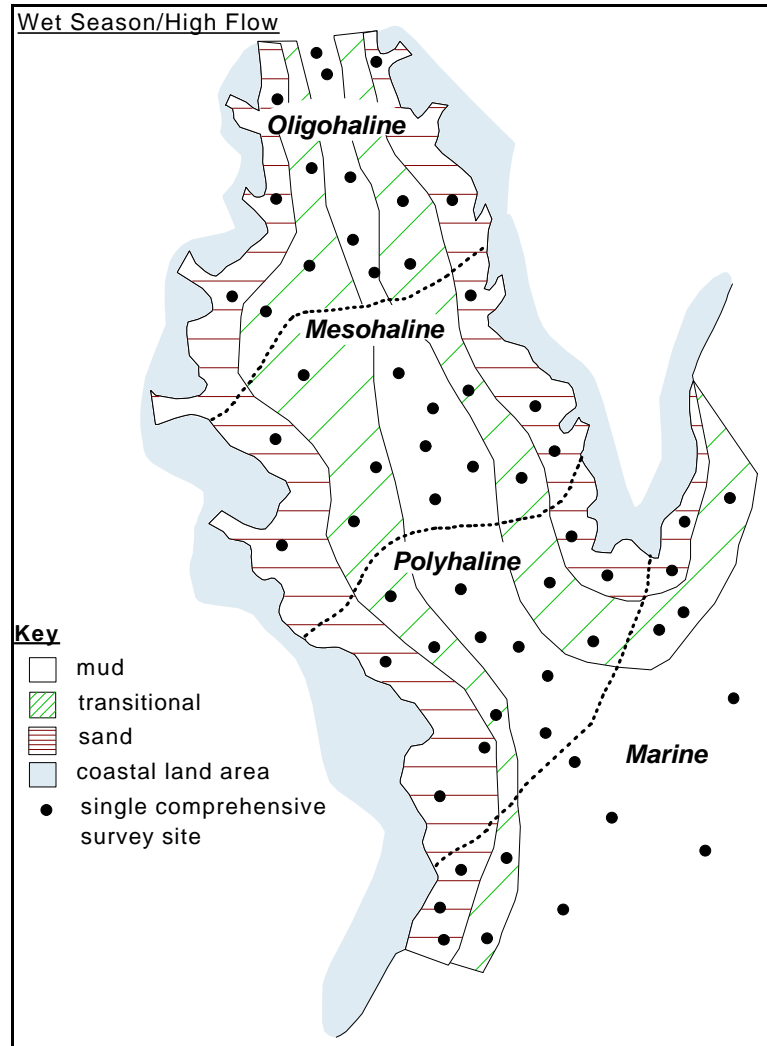
Review of historical data collected in these waters is helpful for establishing potential sample sizes based on the variability in the record. These records are usually available in the published literature, natural history museums, college and university departments, and federal and state agencies. Caution should be exercised in using this information because some biological surveys occurred at impaired sites, may have used incompatible sampling methods, inappropriate or inadequate QA/QC procedures, were insufficiently documented, or had objectives markedly different from biocriteria determination. While important for establishing perspective with respect to current reference site data, historical information alone should not be used to establish precise reference conditions.

4.3.2   Reference Sites

Reference sites refer to locations within a classification category at which data are collected to represent the most natural ambient conditions present. The biocriteria approach generally uses this population of reference sites to establish the collective reference condition that will in turn be used for comparisons of metrics and test sites. Reference sites in estuaries and coastal marine waters include either sites that are distant from point and nonpoint sources and may be applied to a variety of test sites in a given area, or sites that occur along

Physical Classification and the Biological Reference Condition

Table 4-1. Comparison of elements for characterizing reference conditions (adapted from USEPA 1998b).

	Historical Data	Present-Day Biology	Predictive Models	Expert Consensus
Strengths	Yields actual historical information on status. Inexpensive to obtain.	Yields obtainable, best present status. Any assemblages or communities deemed important can be used.	When sufficient data are not available. Work well for water quality.	Relatively inexpensive. Can be better applied to biological assemblages than models. Common sense and experience can be incorporated.
Weaknesses	Data might be limited. Studies likely were designed for different purposes; data might be inappropriate. Human impacts present in historical times were sometimes severe.	Even best sites subject to human impacts. Degraded sites might lower subsequent biocriteria.	Community and ecosystem models not always reliable. Extrapolation beyond known data and relationships is risky. Can be expensive.	May be qualitative descriptions of "ideal" communities. Experts might be biased.

gradients of impact; i.e., nearfield/ farfield.

All monitoring sites, whether reference or test, can vary spatially and temporally due to natural causes. A central measure from several reference sites is used so that natural variability and uncertainty can be accommodated. Statistically, this means that the status of particular estuarine or coastal marine "test" sites are judged by comparing them to a population of reference sites for the particular classification category. There are 3 approaches for using reference sites; these are discussed in Section 4.4.

4.3.3 Models

Mathematical models may be characterized as descriptive or mechanistic. Descriptive models (also known as correlative or statistical models) describe observed relationships among measured attributes of a system. This approach models data without attention to causal factors. Prediction, including forecasting and managing, is the primary goal of a descriptive model, and the model is considered successful if it fits the data well. The utility of descriptive models is often affected by the quantity and quality of data available, and in many cases, insufficient data exists to construct a useful model.

Mechanistic models seek to explain observed relationships as the result of underlying processes - they are also called process models. They typically consist of a set of state variables, which describe how the system is "now", and a set of dynamic equations that describe how the state variables change over time (exogenous variables, or "forcing functions" may also be included). In a sense, mechanistic models are a set of

descriptive models for each component of a system. The objective of mechanistic models is to describe the system itself and not simply the data obtained by taking measurements; i.e., "fitting the data" is not the prime objective. Mechanistic models have many more constraints and are more time-consuming to construct than descriptive models due to the need to match system structure. Despite the fact that these models are not designed for prediction, they are often built and used to forecast and manage ecological resources for the following reasons: 1) in some cases, one does not want to perform an experiment without a reasonable idea of what will happen (e.g. work involving endangered species); 2) some experiments are not feasible - the amount of data needed for a multivariate statistical model grows very rapidly with the number of variables, and obtaining the data required for a descriptive model is prohibitively expensive.

There are two main types of mechanistic models commonly used in biology and ecology. Simulation (also known as management) models are practically oriented and focus on prediction and management. In these models, numerical accuracy is what matters most, the model need not match the system processes and structure. Management models are system specific, resulting in numerical predictions for one particular system. Theoretical (also known as analytical) models focus on scientific understanding of the system. These models are highly analytical, typically involving systems of differential equations, and emphasize principles rather than numerical accuracy. These models have to be simple enough to allow understanding of system behavior and what the model is predicting. This trade-off often requires that the investigator omit or

estimate many quantitative or unknown details, and often assumptions about the interaction of system components represent hypotheses rather than empirically-derived relationships. Theoretical models can apply to many qualitatively similar systems; they are useful whenever the phenomenon of interest occurs across multiple systems.

The degree of complexity of mechanistic models to predict reference conditions is potentially unlimited with attendant increased costs and loss of predictive ability as complexity increases (Peters 1991). However, these models can provide much insight into the interactions which determine ecological condition. Management-oriented mechanistic models sacrifice numerical accuracy in order to capture system dynamics. These models are mathematically complex and require more time and effort to develop than descriptive models. The primary value of mechanistic models may be for understanding ecosystem processes and evaluating likely system responses when mitigation projects are implemented.

4.3.4 Expert Opinion/Consensus

In any data evaluation, it is important to establish a qualified team of regional specialists so the error inherent in professional judgment can be reduced. This team should evaluate the historical data, the candidate reference sites, subsequent data collected, and any models used in the process. This expert team function is even more important when no candidate reference sites are acceptable. Expert consensus then becomes a workable alternative in establishing reference expectations. Under such circumstances, the reference condition may be defined using a consensus of expert opinion based on sound ecological principles applicable to the region of interest.

Three or four biologists are convened for each assemblage to be used in the assessment, and each expert should be familiar with the estuaries or coastal marine waters and assemblages of the region. The experts are asked to develop a description of the assemblage in relatively unimpaired estuaries and coastal marine waters, based on their collective experience. The description developed by consensus will necessarily be more qualitative than quantitative, but metrics and metric scoring can be developed.

It is important that the process used to review the available information and to develop a consensus be thoroughly documented so that it can be repeated in the future if necessary and to provide quality control on its results. This same panel of biologists and natural resource managers may also be consulted in the development of the overall reference condition and subsequent biocriteria. In establishing the team of experts, it should be recognized that bias toward specific assemblages may exist and the team should be appropriately balanced.

4.4 Use of Reference Sites to Characterize Reference Condition

The determination of the biological reference condition from reference sites is based on the premise that estuaries and coastal marine waters least affected by human activity will exhibit biological conditions most natural and attainable for those waters in the region. Anthropogenic effects include all possible human influences, for example, watershed disturbances, habitat alteration (channel dredging and dredged material disposal, shoreline bulkheading), nonpoint source inputs, point source discharges, atmospheric deposition, and fishing pressure. Human activities can be either

detrimental, such as pollutant inputs, or positive, such as responsible resource protection or restoration. In either case, the manager developing a biocriteria program must evaluate the effect of such activities on biological resources and habitat. In practice, most reference sites will have some of these impacts, however, the selection of reference sites is always made from those with the least anthropogenic influences.

Reference sites must be carefully selected because they are used as a key part of the biocriteria benchmark against which test sites are compared. The conditions at reference sites should represent the best range of minimally impaired conditions that can be achieved within a classification category for the region. Two primary considerations guide the selection of reference sites within each site class: minimal impairment and representativeness.

Minimal Impairment - Sites that are relatively undisturbed by human activities are ideal reference sites. However, land use practices and the presence of major urban areas in the basins of many of the nation's estuaries or adjacent to its coastal marine waters have altered the landscape and quality of water resources to such a degree that truly undisturbed sites are rarely available. In fact, it can be argued that no unimpaired sites exist. Therefore, a criterion of "minimally impaired" must be used to determine the selection of reference sites. In regions where minimally impaired sites are still significantly degraded, the search for suitable sites should be extended over a wider area, and multistate cooperation may be essential. It is advisable that the state make every effort, once reference sites are selected, to protect these areas from degradation. This may involve: purchase of land or easements; where

appropriate, location within public reserves; use restrictions or permit constraints on fishing, discharge, or dredging/disposal to protect the quality of the reference area waters.

Representativeness - Reference sites must be representative of the best quality of the estuaries and coastal marine waters under investigation; that is, they must exhibit conditions similar to what would be expected to be found in the region. They should not represent degraded conditions, even if such conditions are the most common. Sites containing locally unusual environmental characteristics can result in uncharacteristic biological conditions and should be avoided.

Once the physical estuarine or coastal marine classification is completed, the biological reference condition should be defined for each class. This can be accomplished with three basic approaches: (1) selected reference sites; (2) determination from population distributions; and (3) site-specific reference sites. The second approach, determination from population distributions, is a relaxation of the requirement for minimal impairment; and the third approach, site-specific reference sites, is a relaxation of the representativeness requirement.

4.4.1 Selected Reference Sites

In this approach, reference conditions are characterized based on the best available sites for a given physical class of estuarine or coastal marine waters, and indexes or models are developed by comparing the best sites (the reference sites) to a second set of sites that may be impaired. The approach assumes that within the population of sites some are minimally disturbed and therefore comprise a minimally impaired biological condition. Selection of

reference sites must be physical or chemical; for example, minimal instances of hypoxia, substantially free of contaminants, a large proportion of natural vegetation in the watershed, little or no industrial point sources, little or no urban runoff, or little or no agricultural nonpoint source pollution. Impaired ("test") sites for testing response of metrics and model building are selected for the presence of one or more such anthropogenic disturbances. Prior definition and selection of reference sites has been used successfully in streams for fish and invertebrate indexes and models (e.g., Barbour et al. 1995, Ohio EPA 1987, Reynoldson and Zarull 1993, USEPA 1987, Wright et al. 1984), and in estuaries for benthic invertebrate indexes (Engle et al. 1994, Summers et al. 1993, Weisberg et al. 1993).

Reference Site Criteria - The overall goal in establishing the reference condition from carefully selected reference sites is to describe the optimal biota that investigators may expect to find at the test sites of interest in the absence of stresses. These "test" or "assessment" sites can then be compared to the reference sites to determine whether impairment exists. The characteristics of appropriate reference sites vary among regions of the country and for different water body and habitat types. In general, the following characteristics (modified from Hughes et al. 1986) are typical of ideal reference sites:

▸ Sediments and water column substantially free of contaminants;

▸ Natural bathymetry, typical of the region;

▸ Natural currents and tidal regime;

▸ Shorelines representative of undisturbed estuaries and coastal

marine areas in the region (generally covered by vegetation with little evidence of shoreline erosion);

▸ Natural color and odor of the water.

In this approach, a single minimally impaired site does not represent any one region or population of sites, and a frequent difficulty is matching habitats for valid comparison, particularly given that the influence of nonpoint source runoff or specific point source discharges may extend over wide areas due to transport of pollutant loads by currents and tides. Reference conditions based on multiple sites are more representative and are important to establishing quantitative-based or numeric biocriteria.

Representative reference sites should be selected within each of the identified classes. A sufficient number of sites are then sampled to adequately characterize the range of existing conditions and to reduce the variability in the measurements for each class. It is desirable to sample a minimum of 10 sites per class, and 30 sites per class is usually optimal for cost effectiveness. In regions where all sites are impacted, the selected number of "best" sites of each class (e.g., mesohaline mud habitat) are sampled, where "best" is determined by least anthropogenic disturbance or impacts, but not by most desirable biota. In regions where the population of minimally impaired reference sites is large, a stratified random sampling scheme (using those sites) will yield an unbiased estimation of reference conditions (Gilbert 1987).

Stressed Sites - Effective metrics respond to environmental degradation and allow discrimination of impaired sites from the reference expectations. Metrics that do not respond are not useful in bioassessment. Response is determined

by sampling a set of stressed sites in the same way as the reference sites. Sites with known problems, such as nutrient loading, thermal pollution, toxic sediments, or those influenced by urban land use, are good candidates. There should be several in each class for adequate tests of metric responses. Since impaired sites are frequently locations of monitoring by water quality agencies, data might already exist to test the biological metrics. However, the sampling methods for reference and impaired sites should be comparable.

For a lengthy sampling season, it is important to account for seasonal shifts of the salinity zone boundaries. Stations proximal to these transition zones may need to be either located far enough away from the boundary to have consistent year-round application or else their classification should be shifted with the seasons. For example, some areas in Figure 4-5 may be polyhaline-sandy bottom in the spring, but in the winter they would be classified as marine-sandy bottom (Figure 4-6). Thus, such stations have a change of classification with the shifting of the halocline. An alternative is to avoid placing stations near the transition zone so that, except in extreme climatic conditions, these stations have consistent habitat characteristics. The biotic data collected at all sites is then subclassified by sediment type (e.g., sand, sandy-mud, mud) and depth for this salinity region. This information becomes the reference condition and part of the biocriteria for any test sites in the region.

Example: EMAP Estuary - The EMAP-Estuaries (EMAP-E) program collected samples in the Virginian and Louisianian provinces. One of the goals of the EMAP-E effort is to develop a statistical benthic index of estuarine condition based on extensive

Figure 4-5

Estuarine and coastal marine biocriteria survey method useful for *a priori* reference site selection. Wet season/high flow salinity pattern showing tributary reference sites and mainstem transects for four salinity and three substrate classifications.

Wet Season/High Flow

Oligohaline

Mesohaline

tributary survey site

area too developed for a reference transect

Polyhaline

NPS loading

Marine

Key

☐ mud

▨ transitional

☰ sand

▨ coastal land area

● single comprehensive survey site

●—● survey site transect across depth profile

🏠 urban areas

🏠 agricultural areas
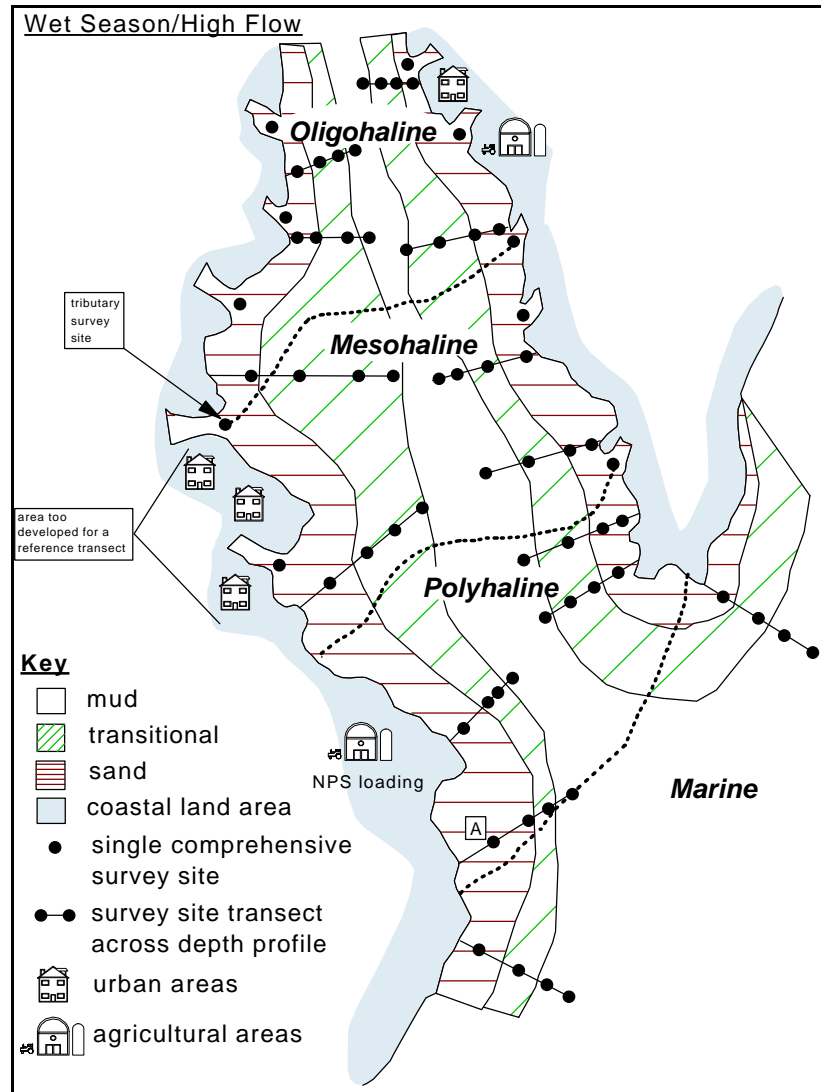
information about benthic community structure. A test data set of reference stations has been compiled for the purpose of formulating the index.

Habitat characteristics used by EMAP to define reference stations from the 1990 and 1991 Virginian province (refer to Figure 4-3) collections in Chesapeake Bay were:

▸ Stations where no contaminant exceeded the effects range-median (ER-M) value (which equals the concentration at which 50% of collected data demonstrated adverse biological effects [Long et al. 1995]);

▸ No sediment toxicity was observed; i.e., percent survival greater than 75% and not significantly different from controls;

▸ Bottom DO was never less than 1-mgL^{-1}, 90% of the continuous DO measurements were greater than 3-mgL^{-1} and 75% of the DO measurements were greater than 4-mgL^{-1} (Schimmel et al. 1994).

The list of stations generated using these characteristics was reviewed to eliminate any reference sites located in areas potentially subject to physical disturbance, such as dredged shipping channels. Fifty-three sites from the
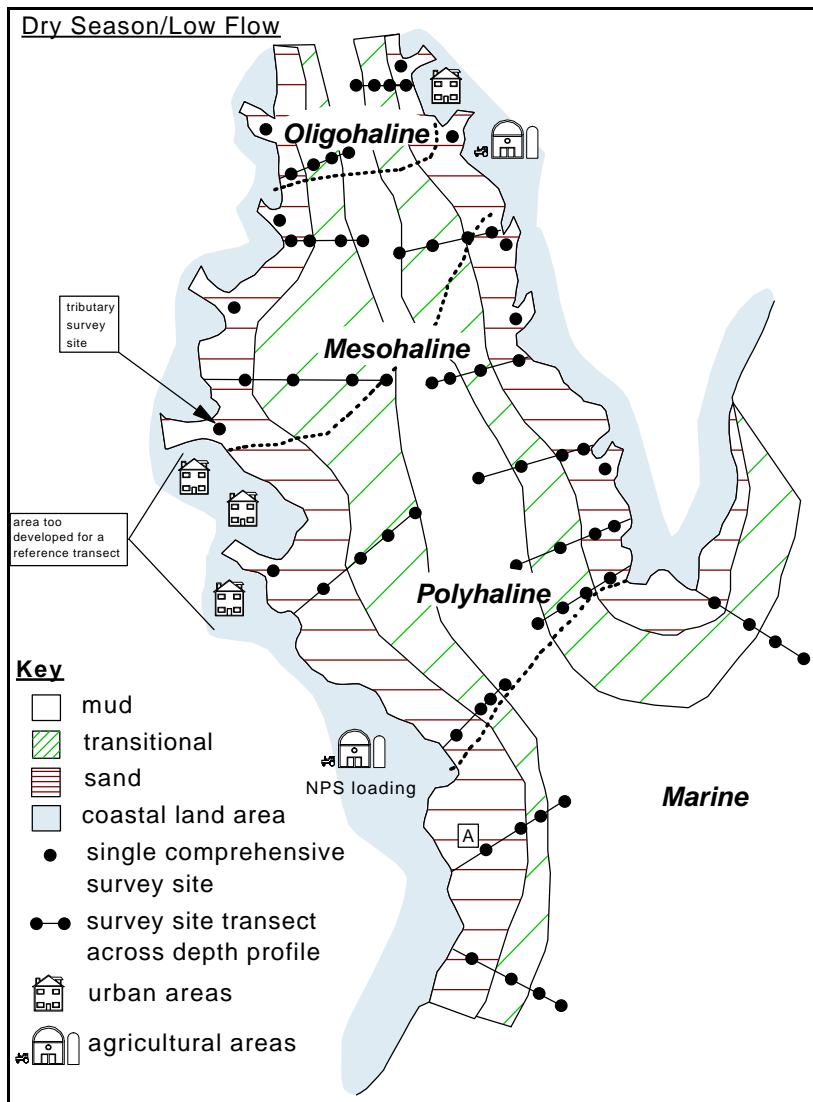
Figure 4-6

Estuarine and coastal marine biocriteria survey method useful for *a priori* reference site selection. Dry season/low flow salinity pattern showing tributary reference sites and mainstem transects for four salinity and three substrate classifications.

combined 1990 and 1991 data sets were considered to be reference sites.

A similar process has been used for data collected in 1991 in the Louisianian province (refer to Figure 4-3). Using the following criteria, eight sites were classified as reference sites:

▸ The minimum DO value over a 24-hour period was less than 3.0-mgL^{-1} (Summers and Engle 1993);

▸ Sediment concentrations for any contaminant did not exceed the ER-M value;

▸ The percent survival for *Ampelisca abdita* (10-day) or *Mysidopsis bahia* (96-hour) in acute sediment bioassays was indistinguishable from controls (Engle et al. 1994).

As states develop their estuarine and coastal marine biocriteria, they may wish to consider incorporating EMAP-identified reference sites into their sampling programs. To the degree that these stations meet state reference condition requirements, they can serve as regional reference sites within the appropriate state classification categories while also contributing to USEPA national trend monitoring for estuaries.

4.4.2 Reference Condition Derived from Population Distribution

One problem in the use of the minimally impaired sites technique is what to do if an area is so extensively degraded that even the least impaired site indicates significant deterioration. Many systems are greatly altered through channel dredging and spoil disposal, urbanization, and construction and operation of marinas and other commercial or industrial enterprises. The condition of these systems is a result of societal decisions that have to be taken into account. However, the existence of greatly altered systems should not compromise the objective of defining the natural state as a reference condition. These disturbed systems should not be presumed to represent a reference condition of any sort.

Although the biocriteria established for these altered systems serve as a baseline for judging impairment, the ultimate goal is to achieve the sites' recovery to the best attainable condition as represented by historical information and by conditions at "minimally impaired" sites. Consensus of expert opinion and historical data play an especially important role in characterizing the reference condition for these systems, as does the application of innovative management practices to obtain resource improvement.

In defining the biocriteria, managers must strike a balance between the ideal restoration of the water resource and the fact that human activity affects the environment. The most appropriate course of action will be to use **minimally** impaired sites as representing the **maximum** amount of degradation that will be tolerated, thereby ensuring adherence to the antidegradation policy of the CWA.

Continual monitoring should provide the feedback necessary to make reference condition and interim criteria adjustments as warranted during the restoration process.

In this approach, reference conditions are derived from the distribution of calculated metrics for the entire biological data set within a physical classification without preselecting any reference sites. The entire data set can be plotted as a cumulative frequency distribution to help determine "best" values of candidate metrics (Figure 4-7). This approach is applied in cases where prior definition of reference sites is not possible because all sites are considered impaired or because too few reference sites exist (e.g., one or two) for an unbiased characterization of regional reference conditions. This approach has been used successfully for fish and invertebrate indexes in streams (e.g., Karr et al. 1986, Plafkin et al. 1989) and for fish (Jordan et al. 1992, Deegan et al. 1997) in estuaries.

The biological reference condition is defined from some upper fraction of the component indicator variables (metrics) and this reference condition is subsequently used to judge the biological status of other sites. There is no independent (nonbiological) definition of reference condition. Reference condition and biological responses are confirmed by identifying severely impaired sites and then comparing them with the derived reference condition to determine the response(s) of biological indicators to impacts, and by selecting metrics that are known to respond to perturbation from other studies.

A representative sample is taken of the entire population of estuary or coastal marine sites (Figure 4-8). Sites that are known to be severely impaired may be

Physical Classification and the Biological Reference Condition
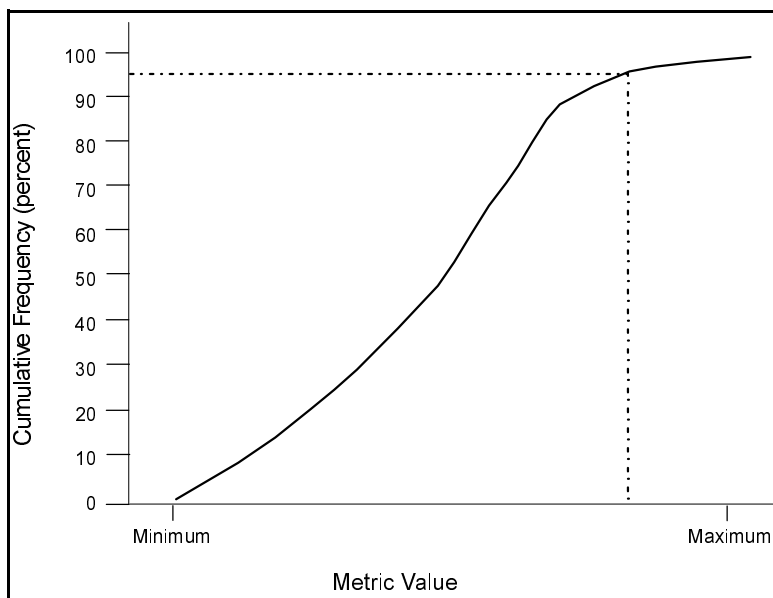
Figure 4-7

Hypothetical cumulative frequency distribution of metric values for all sites in a given estuarine or coastal marine class. The dotted line shows the metric value corresponding to the 95th percentile.

excluded from the sample, if desired. The population distribution of each biological metric (Chapter 11) is determined, and the 95th percentile of each metric is taken as its reference value. The range from the minimum possible value to the reference value is trisected, and values in the top third of the trisected range are presumed to be similar to reference conditions. Scoring of metrics is explained more fully in Chapter 11.

A central assumption of the population distribution approach is that at least some sites in the population of sites are in good condition, which will be reflected in the highest scores of the individual metrics. Because there is no independent definition of reference; i.e., independent of biological status, reference conditions defined in this way must be taken as interim and subject to future reinterpretation. Again, antidegradation safeguards must be in place to prevent further deterioration of the reference condition and criteria.

4.4.3 Site-specific Reference Sites

The site-specific approach is analogous to upstream-downstream comparisons

in running water or control-impact designs. It consists of selecting a reference site paired with each site to be assessed. There is no characterization of reference conditions for a physical class of estuarine or coastal marine waters; each test site and each reference site is a special case with each test site compared to its reference site. Reference sites are selected to be similar to their respective test site, but unimpaired by the perturbations of interest at the test site. This approach may be less costly at the outset because the design and logistics are simpler than the other approaches. However, after several years of sampling and monitoring, costs for this approach are likely to be similar or greater because each new test site requires its own paired reference site.

The site-specific approach has two problems stemming from the fact that there is usually only a single reference site or a single nearby reference area from which reference sites are selected. The first problem is representativeness: Does the reference site represent reference conditions? Although the reference site may lack the specific stressor that is present at the test site, unless carefully evaluated and placed, it

Figure 4-8

Estuarine and coastal marine biocriteria survey method useful for stratified random (population distribution) reference site selection. Dry season/low flow salinity pattern showing mainstem sampling sites for four salinity and three substrate classifications.
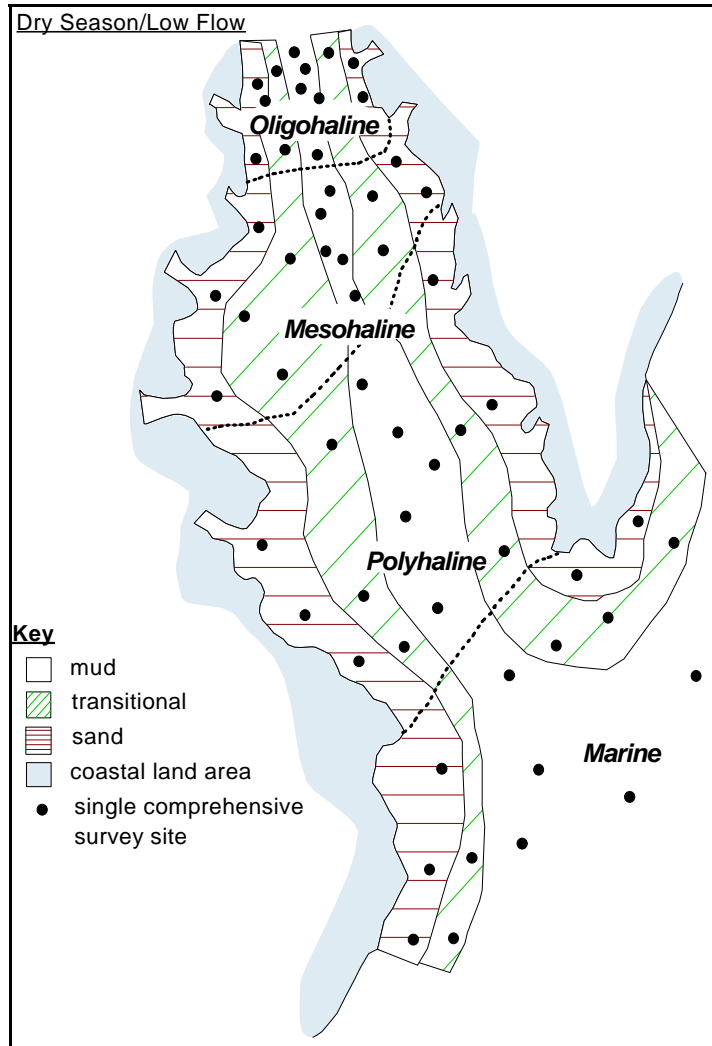
may be subject to other stressors that have not been considered.

The second problem with the site-specific approach is the potential for trivial statistical comparison of two sites in that it is almost always possible to demonstrate a statistically significant difference between two sites by pseudoreplication (Hurlbert 1984). Pseudoreplication is the repeated measurement of a single experimental unit or sampling unit, and treating the measurements as if they were independent replicates of the sampling unit. A single reference site does not yield sufficient information to meaningfully judge the biological relevance of a statistical difference at the test site. The judgment that biotic

differences between a single test site and its reference site may be due to differences in impacts can not depend on statistical tests, but requires a careful weight-of-evidence evaluation (e.g., Hurlbert 1984, Schindler 1971).

If the objective of a study is to test the response of a particular metric, and if there are several paired sites, then a paired approach can be very powerful, allowing paired statistical tests (e.g., Frydenborg 1994). A paired experimental design is not pseudoreplication because each site pair is an independent replicate, and the sample size (n) is the number of pairs.

Example 1: Navigation channels -

Navigation channels can represent an important component of overall estuarine areas (e.g., Houston Ship Channel, entrance to Chesapeake Bay and major harbors). Resource agencies may need to determine the relative quality of navigation channels in relation to the entire estuarine system as part of the overall resource evaluation.

Stations should be arrayed essentially in a nearfield-farfield pattern as shown in Figure 4-9, with farfield stations located "up" current and nearfield stations "down" current, outside the zone of suspected impact. Stations should be located such that depth, grain size, and salinity remain consistent. These conditions may be difficult to locate in a tidally-influenced channel. Furthermore, if the navigation channel to be assessed is dredged to constant depth, changes in biota will primarily be a function of salinity, given uniform poor substrate and the periodic destruction of the benthic habitat by dredging.

The reference condition for navigation channels would be determined from the central tendency (e.g., median) of the biological data collected at "upstream" stations, that is, those stations that are expected to be out of the zone of influence of impact sources (e.g., harbors, industrial areas). Sites from which the reference condition is determined should be of comparable depth, grain size, and salinity to those in suspected impact zones and have the same dredging history.

Example 2: Nearshore marine -

The station array in coastal marine waters is essentially a variation of the nearfield-farfield approach because of the open water characteristics. Transects should be laid parallel to shore along equal depth contours, with sampling stations placed approximately evenly along the transect (Figure 4-10). For habitat consistency, the survey team should strive to maintain uniform depth, bottom type, salinity, DO, and pH characteristics at a minimum for all sites.

These parallel transects can evolve to an open grid station array if sampling stations are added around outfalls. In Figure 4-10, the D1-D5 series of stations is added to the transect to reveal effluent distribution shifts around the discharge site. This approach addresses two aspects of effluent impact monitoring: (1) the relative biological community change near the discharge as compared to the reference condition described by observation of either end of the transect; and (2) the potential shifting, seasonal change, or expansion or contraction of the zone of effluent influence from the discharge. Both forms of information are important to adequately assess the biological effects of such effluents. This design was used for bioassessment of ocean outfalls from Delaware and Maryland (see Chapter 13).

A complete grid, while more involved and expensive, would allow a more precise evaluation of the effects of these discharge plumes as they shift in position in response to changes in nearshore currents and seasonal shifts in wave regime (Figure 4-10).

Figure 4-9

Estuarine and coastal marine survey method for navigation channel assessment.

"Up" channel stations

Harbor

Impact Zone

Net Current Direction

"Down" channel stations

Key
● Test sites
■ Reference sites
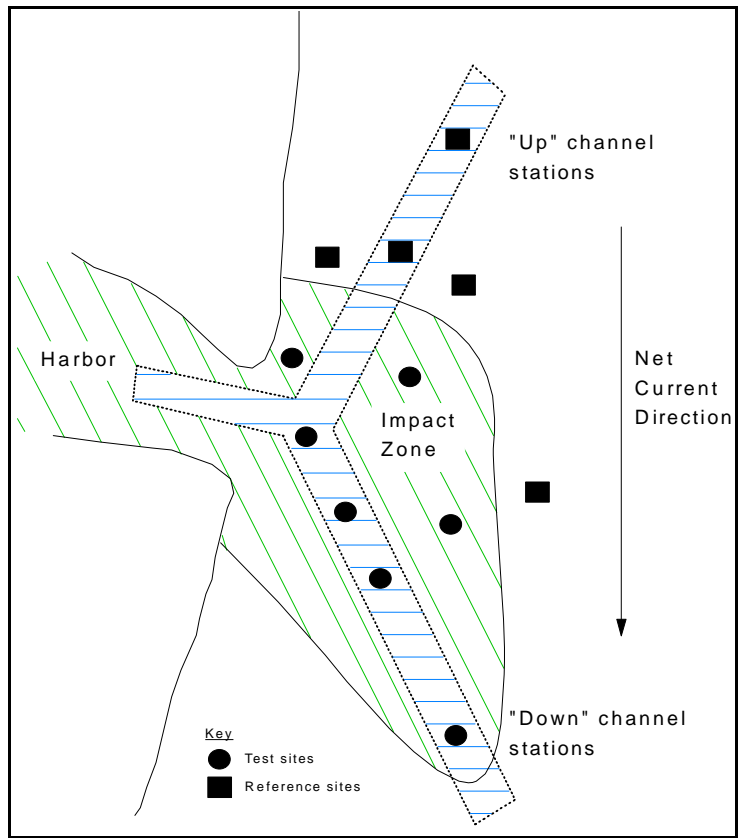
Figure 4-10

Estuarine and coastal marine biocriteria survey method useful for marine site selection.

net longshore current

● A
● B
● C
D1 ● ● D3
● D ● D4
D2 ● ● D5
● E
● F
● G
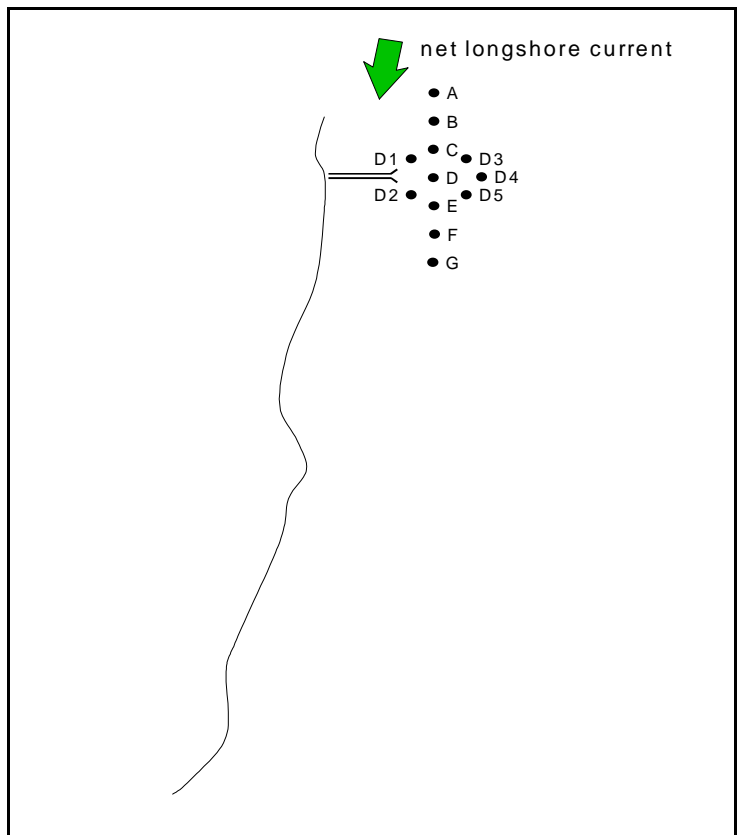
Chapter 5

Sampling Program Issues Biological Assemblages and Design

This chapter presents sampling program issues that are common to each of the three assessment tiers that employ field sampling. These issues include the biological assemblages (Section 5.1) that might be sampled, sampling design strategies (Section 5.2 and 5.3), and logistical considerations (Section 5.4). Historically, benthic macroinvertebrates have been the most widely sampled assemblage, which is described in detail in this chapter.

As described earlier in this document, a possible sampling methodology is a progressive tiered design, ranging from simple biological assessment to detailed, intensive studies. The tiers are intended to be implemented cumulatively, that is when possible, each tier should incorporate the elements in the preceding tier as appropriate for the estuaries or coastal marine water in which they are applied. In general, the methods are derived from those used along the coastal United States (Dauer 1993, Farrell 1993a, b, Nelson et al. 1993, Word 1980, 1978, Word et al. 1976); in Puget Sound (Eaton and Dinnel 1993); in the EMAP - Estuaries program (Holland 1990), and in USEPA's National Estuary Program (NEP) (USEPA 1992) and 403 Monitoring Program (USEPA 1994a).

Assessment tiers 1 through 3 require sampling biological assemblages and habitats in one or more field visits. Six biological assemblages, including two developmental/experimental assemblages, are recommended for estuarine and coastal marine waters

bioassessment. Each tier is comprised of a subset of assemblages, with the number of assemblages increasing in the higher tiers. While these six assemblages are described, specific environmental circumstances and budget constraints will determine what subset each state uses. For example, if finances are extremely limited on the East Coast the single most effective assemblage to sample may be macrophytes. On the West Coast benthic macroinvertebrates or fish may the assemblages of choice. The bioassessment measurements are made along transects extending from shore to the deepest (channel) portion of the estuary, in a systematic grid along transects extending away from point source discharges (nearfield/farfield), or in a probablistic design. The number of transects or grid points, the assemblages sampled, and the intensity of sampling effort are determined by the assessment tier with overall effort increasing at each higher tier.

5.1 Assemblages

The study of any group of organisms will yield information on the status of their environment. The objectives in selecting assemblages for estuarine and coastal marine bioassessment were to identify those that: (1) are unambiguously useful for biological assessment; (2) can be sampled and interpreted in a cost-effective way; and (3) have easily calculated metrics that can be used alone or in a multimetric index of the assemblage. Assemblages

that meet these criteria are suggested for use in estuarine and coastal marine assessment; assemblages that do not presently meet the criteria are considered to be developmental. Suggested assemblages include infaunal benthic macroinvertebrates, fish, aquatic macrophytes, and phytoplankton (chlorophyll *a*). The developmental assemblages include zooplankton, epibenthos, and paleoenvironmental systems. These developmental assemblages are promising, but they lack the same level of refinement documented for the suggested assemblages listed above and unresolved technical problems remain with respect to cost-effective assessment and interpretation. Background and rationale for these suggested assemblages was presented in Chapter 2.

Multimetric bioassessment is not a ready-made, one-size-fits-all instrument that will tell managers whether estuaries or coastal marine waters are healthy. It is an approach that is expected to be modified to specific regional conditions before it can be applied. For example, bioassessment of streams has been successful when modified and calibrated regionally (e.g., Barbour et al. 1996a, Ohio EPA 1990, Miller et al. 1988), but it has been less successful when used "off-the-shelf." Successful application requires region-specific selection and calibration of metrics, as well as regional characterization of reference conditions. For example, benthic infauna are rare in rocky, fjord-type estuaries and would be an inappropriate assemblage to sample in such a setting.

5.1.1 Benthic Macroinvertebrates (Infauna)

Benthic macroinvertebrates are an appropriate assemblage for all biological assessments of water bodies because they respond to water, sediment, and

habitat qualities (Holland 1990, Plafkin et al. 1989), are not very mobile, and consequently, integrate long-term changes in these ecosystem components. For those reasons, benthic macroinvertebrates tend to dominate this text.

Individual macroinvertebrate species have sensitive life stages that respond to stress and integrate effects of short-term environmental variations, whereas community composition depends on long-term environmental conditions. In addition to taxonomic identification, benthic macroinvertebrate metrics may require knowledge of the feeding group to which a species belongs, for example, suspension feeders and deposit feeders. Potential metrics for estuarine and coastal marine benthos are listed in Table 5-1. Metrics considered in the EMAP Estuaries program are listed in Table 5-2.

Sampling Strategies

The sampling area should focus on the most predominant substrate available (in many estuaries and coastal marine areas this will be soft sediments of mud through sand grain sizes), and the metrics should be developed independent of microhabitat variation (Table 5-3). The type of sampling gear will depend on the substrate being sampled; each substrate has its own optimal sampling gear (Section 5.1.1.4). Standardized sampling techniques for each gear type should be followed to allow for the comparison of data. Processing of samples should be standardized by using a mesh size appropriate to the region. In the past, monitoring programs conducted in east coast waters have often used a 0.5-mm mesh screen, while west coast programs have used a 1.0-mm screen (Bowman et al. 1993). States should consider testing various mesh size screens to determine

Table 5-1.   Potential benthic macroinvertebrate metrics.

Metric	Response to Impairment
No. of taxa	reduced
Mean no. of individuals per taxon	substantially lower or higher
% contribution of dominant taxon	elevated
Shannon-Wiener diversity	reduced
Total biomass	substantially lower or higher
% biomass of opportunistic species	elevated
% abundance of opportunistic species	elevated
Equilibrium species biomass	reduced
Equilibrium species abundance	reduced
% taxa below 5-cm	reduced
% biomass below 5-cm	reduced
% carnivores and omnivores	elevated
No. of amphipod species	reduced
% individuals as amphipods	reduced
% individuals as polychaetes/oligochaetes	elevated
No. of bivalve species	reduced
% individuals as molluscs	reduced
% individuals as deposit feeders	elevated
Mean size of organism in habitat	reduced
Proportion of expected no. of species in sample	reduced
Proportion of expected no. of species at site	reduced
Mean weight per individual polychaete	reduced
No. of suspension feeders	reduced
% individuals as suspension feeders	reduced
No. of gastropod species	reduced
No. of Capitellid polychaete species	elevated

the most appropriate size for their bioassessment activities. Ferraro et al. (1994) present a process to evaluate the optimum infaunal sampling protocol; i.e., sampling unit area, sieve mesh size, and sample size [*n*], discussed more fully in Section 5.2.6.

Time and Costs

An informal survey of some states that conduct routine monitoring of estuaries and coastal marine waters indicates that estuarine sampling requires a minimum of two full-time equivalent (FTE) staff,

Table 5-2.   Metrics from which the EMAP Virginian and Louisianian benthic indexes were developed. Louisianian Province has reduced number of metrics due to knowledge gained from previous Virginian province studies (n.a. - not applicable).

Community Measure of Structure/ Function	Metrics
Virginian Province	
Biodiversity/ Species Richness	Proportion of expected number of species present in a sample ■ Proportion of expected number of species present at a site ■ Shannon-Weiner Diversity Index ■ Pielou's evenness index
Abundance Measures	Total benthic abundance per event ■ Mean benthic abundance per sample ■ Total benthic biomass per event ■ Mean benthic biomass per sample
Individual Health	Biomass/abundance ratio ■ Mean weight per individual polychaete ■ Mean weight per individual mollusc
Functional Groups	Number of suspension feeding organisms per event ■ Biomass of suspension feeding organisms per event ■ Percent of total benthic abundance as suspension feeders ■ Percent of total benthic abundance as suspension feeding biomass ■ Number of deposit feeding organisms per event ■ Biomass of deposit feeding organisms per event ■ Percent of total benthic abundance as deposit feeding organisms ■ Number of benthic omnivores/predators per event ■ Biomass of benthic omnivores/predators per event ■ Percent of total benthic abundance as omnivores/predators ■ Percent of total benthic biomass as omnivores/predators ■ Number of opportunistic species per event ■ Mean number of opportunistic species per sample ■ Percent of total benthic abundance as opportunists ■ Number of equilibrium species per event ■ Mean number of equilibrium species per sample ■ Percent of total benthic abundance as equilibrium species ■ Percent of mean benthic abundance as equilibrium species
Taxonomic Composition	Number of amphipods per event ■ Amphipod biomass per event ■ Percent of total benthic abundance as amphipods ■ Percent of total benthic biomass as amphipods ■ Number of bivalves per event ■ Bivalve biomass per event ■ Percent of total benthic abundance as bivalves ■ Percent of total benthic biomass as bivalves ■ Number of gastropods per event ■ Gastropod biomass per event ■ Percent of total benthic abundance as gastropods ■ Percent of total benthic biomass as gastropods ■ Number of molluscs per event ■ Mollusc biomass per event ■ Percent of total benthic abundance as molluscs ■ Percent of total benthic biomass as molluscs ■ Number of polychaetes per event ■ Polychaete biomass per event ■ Percent of total benthic abundance as polychaetes ■ Percent of total benthic biomass as polychaetes ■ Number of Capitellid polychaetes per event ■ Percent of total benthic abundance as Capitellid polychaetes ■ Number of Spionid polychaetes per event ■ Percent of total benthic abundance as Spionid polychaetes ■ Percent of total polychaete abundance as Spionid polychaetes ■ Number of Tubificid oligochaetes per event ■ Percent of total benthic abundance as Tubificid oligochaetes
Louisianian Province	
Biodiversity/ Species Richness	Shannon-Wiener Diversity Index ■ Pielou's Evenness Index ■ Mean number of species ■ Mean number of polychaete species
Abundance Measures	Mean benthic abundance per site
Individual Health	n.a.
Taxonomic Composition	Mean abundance of amphipods per site ■ Proportion of total benthic abundance as amphipods ■ Mean abundance of decapods per site ■ Proportion of total benthic abundance as decapods ■ Mean abundance of bivalves per site ■ Proportion of total benthic abundance as bivalves ■ Mean abundance of gastropods per site ■ Proportion of total benthic abundance as gastropods ■ Mean abundance of molluscs per site ■ Proportion of total benthic abundance as molluscs ■ Mean abundance of polychaetes per site ■ Proportion of total benthic abundance as polychaetes ■ Mean abundance of Capitellid polychaetes per site ■ Proportion of total benthic abundance as Capitellid polychaetes ■ Mean abundance of Spionid polychaetes per site ■ Proportion of total benthic abundance as Spionid polychaetes ■ Proportion of total polychaete abundance as Spionid polychaetes ■ Mean abundance of Tubificid oligochaetes per site ■ Proportion of total benthic abundance as Tubificid oligochaetes

Table 5-3. Sampling summary for infaunal benthic macroinvertebrates.

Habitat	Preferred: soft sediments (mud-sand).
Sampling Gear	Regionally most appropriate for substrate (Table 5-4).

Index Period	Regionally most appropriate
	Preferred:
	      Summer - East & Gulf Coast
	      Spring - Pacific Northwest
	Alternative:
	      All four seasons, or winter and summer

Sampling	Preferred: samples from 3 grabs at each of at least 10 sites. Alternative: keep sites as replicates if a within-class variance estimate will be used in assessment.
Analysis	Preferred: lowest practical taxonomic level Alternative: identification to class and family.

and has an associated per sample cost of $200 - $400.

Coastal marine sampling requires a minimum of four FTEs, and has an associated cost of $400 - $800. Three months to a year are required from time of sampling to preparation of an interpretive report.

Assessment Tiers

The benthic infaunal assemblage is appropriate for all three field tiers outlined for the biological assessment of estuaries and coastal marine waters. Tier 1 determines the presence/absence of macroinvertebrates below 5-cm depth in the sediment and briefly describes the class and family of observed benthos. Tier 2 determines the major taxa and indicator species present in each sample to the genus and species level. Tier 3 applies a full benthic community assessment, recording the numbers of individuals in each grab to the genus and species level, and can include determination of biomass if deemed appropriate by the state. Tier 3 uses the benthic community assessment with replication and additional diagnostic

stations and parameters as indicated by the data.

Gear Type

All sampling methods and gear types have specific biases because they capture a target assemblage. Because estuaries and coastal marine waters are complex environments with a potentially large number of habitats, it is important to choose sampling methods and gear appropriate for a specific habitat type. Sampling within a given habitat type such as a salinity regime, bottom grain size, and/or depth should be conducted so that samples can be considered representative of the community being studied.

A large number of benthic sampling methods and gear are available. The choice of appropriate methods and gear will depend upon the goals of the sampling and the habitat to be sampled.

‣ In subtidal areas, benthic infauna can be collected using grabs, such as Young, Ponar, or Van Veen; or cores such as box, gravity, or hand-held cores collected by divers. Grab or

core size and number of replicates should be sufficient to adequately sample the infaunal community, bearing in mind that distribution is usually spatially clumped rather than random or regular; and

▸ Intertidal areas may best be sampled at low tide with hand-held cores. For certain infauna it may also be feasible to estimate abundance by counting the number of surface structures within a given area. For example, some polychaete worms build identifiable tube or mound structures, or leave identifiable fecal coils in intertidal areas. If the local infauna has been studied to the extent that identification of such topographic features can be correlated to the presence of a particular organism, crude abundance and presence/absence evaluations may be possible.

Collection of sediments and benthic organisms should be done concurrently in order to reduce the costs of field sampling and to permit sound correlation and multivariate analyses. Therefore, the sampling equipment and procedure should also include sampling the sediment.

Desirable attributes for sediment sampling gear include:

▸ Creates a minimal pressure wave when descending;

▸ Forms a nearly leakproof seal when the sediment sample is taken;

▸ Prevents winnowing and excessive sample disturbance when ascending;

▸ Allows easy access to the sample surface so that undisturbed subsamples may be taken;

▸ Allows vertical sectioning of undisturbed samples for profile examination.

Penetration well below the desired sampling depth is preferred to prevent sample disturbance as the device closes. It is best to use a sampler that has a means of weight adjustment so that penetration depths may be modified with changing sediment type (USEPA 1992).

Grab Samplers

Well designed and constructed grab samplers are capable of consistently sampling bottom habitats. Depending on the size of the device, areas of 0.02- to 0.5-m^2 and depths ranging from 5- to 15-cm may be sampled. Limitations of grab samplers include:

▸ Variability among samples in penetration depth depending on sediment properties;

▸ Oblique angles of penetration which result in varying penetration depths within a sample; and

▸ The sample may be folded or otherwise distributed by some devices, such as the Shipek sampler, resulting in the loss of information concerning the vertical structure of benthic communities in the sediments.

However, careful use of these devices will provide reliable quantitative data. Grab samplers are the tools of choice for a number of estuarine and marine monitoring programs due to their ability to provide quantitative data at a relatively low cost (Fredette et al. 1989, USEPA 1986-1991). Various grab samplers which could be used for Tiers 1-3 are summarized in Table 5-4.

Table 5-4. Summary of bottom sampling equipment (Adapted from USEPA 1992, Klemm et al. 1992, and ASTM 1998b).

DEVICE	USE	ADVANTAGES	DISADVANTAGES
KB Corer	Soft sediments only.	Samples a variety of soft substrates up to harder types. Sampling tube can be modified up to 100-cm² substrate surface; least disturbance to water/bottom interface. Can be used in shallow to medium-shallow water up to 30.5-m or deeper.	Samples limited surface area. Requires boat and winch.
Ballcheck Single and Multiple Tube	Soft sediment only.	Good penetration on soft sediment. Small sample volume allows greater number of replicates to be collected in a short time period. Samples deep burrowing organisms. Used in shallow to deep water (3-m to 183-m). Automatic check valves prevent sample loss.	Heavy; requires boat and winch. Does not retain sand unless bronze core retainers are used.
Fluorocarbon plastic or Glass Tube	Shallow wadeable waters or deep waters if SCUBA available. Soft or semi-consolidated deposits	Preserves layering and permits historical study of sediment deposition. Rapid-samples immediately ready for laboratory shipment. Minimal risk of contamination.	Small sample size requires repetitive sampling. Impractical in water > 1-m depth if SCUBA not available.
Hand Corer with removable Fluorocarbon plastic or glass liners.	Same as above except more consolidated sediments can be obtained.	Handles provide for greater ease of substrate penetration. Above advantages.	Careful handling necessary to prevent spillage. Requires removal of liners before repetitive sampling. Slight risk of metal contamination from barrel and core cutter.
Box Corer	Same as above.	Collection of large undisturbed sample allowing for subsampling.	Hard to handle.
Phleger (Gravity) Corer	Semi-consolidated sediments.	Low risk of sample contamination. Maintains sediment integrity relatively well.	Careful handling necessary to avoid sediment spillage. Small sample, requires repetitive operation and removal of liners. Time consuming.
Young Grab	Lakes, estuarine and marine areas.	Eliminates metal contamination if grab is plastic or kynar lined. Reduced pressure wave. Can subsample. Better penetration in sand than the modified Van Veen.	Expensive, heavy, requires boat and winch.
Ekman or Box Dredge	Soft to semi-soft sediments. Can be used from boat, bridge, or pier in waters of various depths. Weights can be added for deeper penetration in fine sand.	Obtains a larger sample than coring tubes. Can be subsampled through box lid. Hinged top doors reduce washout, shock waves and substrate disturbance. Range of sizes available.	Possible incomplete jaw closure and sample loss. Possible shock wave which may disturb the fines. Metal construction may introduce contaminants. Possible loss of fines on retrieval. Inefficient in deep water or where even moderate current exists.

Table 5-4 (Cont'd). Summary of bottom sampling equipment (Adapted from USEPA 1992, Klemm et al. 1992, and ASTM 1998b).

DEVICE	USE	ADVANTAGES	DISADVANTAGES
Ponar Grab Sampler	Useful on sand, silt, or clay.	Most universal grab sampler. Adequate on most substrates; very efficient for hard sediments. Large sample obtained intact permitting subsampling. Better penetration than other grabs; sideplates and screens reduce washout, shock waves and substrate disturbance.	Shock wave from descent may disturb fines. Possible incomplete closure of jaws results in sample loss. Possible contamination from metal frame construction. Sample must be further prepared for analysis. A very heavy grab requires use of a boat with winch and cable. Shell hash can hold jaws open causing loss of sample. Must use stainless-lined grab for sediment metals samples.
BMH-53 Piston Cover	Waters of 1-2-m deep when used with extension rod. Soft to semi-consolidated deposits.	Piston provides for greater sample retention.	Cores must be extruded on site to other containers - metal barrels introduce risk of metal contamination.
Modified Van Veen	Useful on sand, silt, or clay.	Adequate on most substrates. Large sample obtained intact.	Requires boat and winch. Shock wave from descent may disturb fines. Possible incomplete closure of jaws results in sample loss. Possible contamination from metal frame construction. Sample must be further prepared for analysis. Limited penetration in hard sand. Possible overpenetration in soft silt.
BMH-60	Sampling moving waters from a fixed platform.	Streamlined configuration allows sampling where other devices could not achieve proper orientation.	Possible contamination from metal construction. Subsampling difficult. Not effective for sampling fine sediments.
Smith-McIntyre Grab	Useful on most substrates.	Reduced pressure wave. Designed for sampling hard substrates. Can subsample and make vertical cross-sections. Greater penetration in sand and cobble than modified Van Veen, but possibly not as deep as a Young grab. Better closure in areas with wood debris.	Loss of fines. Heavy; requires boat and winch. Possible metal contamination unless grab is lined.
Scoops, Drag Buckets	Useful on most substrates. Various environments depending on depth and substrate.	Inexpensive, easy to handle.	Loss of fines on retrieval through water column. Layer information not collected.

The number and kinds of macroinvertebrates collected by a particular grab may be affected by the habitat sampled, substrate type sampled, depth of penetration, angle of closure, completeness of closure of the jaws and potential loss of sample material during retrieval, creation of a "shock" wave and "washout" of organisms at the surface of the substrate. The high-flow velocities often encountered in rivers and wave action in estuaries and coastal marine waters can also affect stability of the sampler (Klemm et al. 1992). USEPA EMAP-Estuaries protocols describe a simple and consistent method for accepting or rejecting a bottom grab (Figure 5-1).

The type and size of the grab samples (or other device) selected for use will depend on factors such as the size of boat, available winch and hoisting gear, the type of sediment to be sampled, water depth, current velocity, and whether sampling is conducted in sheltered areas or open water (Klemm et al. 1992). The EMAP-Near Coastal

Figure 5-1

Cross-section of sediment in clamshell bucket illustrating acceptable and unacceptable grabs.

Acceptable if Minimum Penetration Requirement Met and Overlying Water is Present

Unacceptable (Washed, Rock Caught in Jaws)

Unacceptable (Canted with Partial Sample)

Unacceptable (Washed)
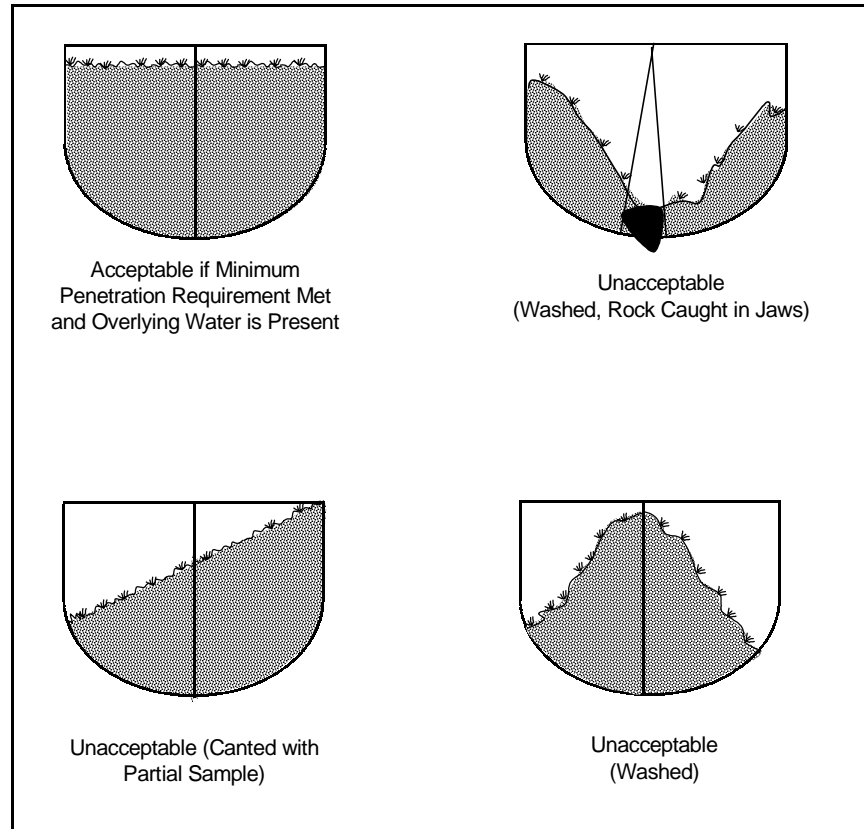
Program selected a Young grab (sometimes referred to as a Young-modified Van Veen) that samples a surface area of 440-cm^2 (Weisberg et al. 1993). This Young grab was selected because it deploys easily from small boats (24-ft) and it samples sand and mud habitats adequately. The maximum penetration depth of the grab was 10-cm.

PONAR Grab:

The PONAR has side plates and a screen on the top of the sample compartment to prevent loss of the sample during closure. With one set of weights, this heavy steel sampler can weigh 20-kg. Word et al. (1976) report that the large amount of surface disturbance associated with Ponar grabs can be greatly reduced by simply installing hinges rather than fixed screen tops, which will reduce the pressure wave associated with the sampler's descent

into the sediment. The standard Ponar takes a sample area of 523-cm^2. A small version, the petite Ponar grab, takes a sample area of 232-cm^2 and can be used in habitats where there may be an unusual abundance of macroinvertebrates, thus eliminating the need to subsample.

The weight of the standard Ponar grab makes it necessary to use a winch and cable or portable crane for retrieving the sample, and ideally the samples should be taken from a stationary boat. The smaller version (petite Ponar grab) is designed for hand-line operation, but it may be used with a winch and cable.

Ekman Grab:

The Ekman grab sampler is used to obtain samples of macroinvertebrates from soft sediments, such as very fine sand, mud, silt, and sludge where there is little current. This grab is inefficient

in deep waters, under adverse weather conditions, and in waters with moderate to strong currents or wave action. The Wildco box corer is like a heavy duty Ekman with a frame and weights and can be used to collect macroinvertebrates in estuaries. Because of its weight a winch is necessary for retrieving the sample from a stationary boat.

The Ekman grab sampler is a box-shaped device with two scoop-like jaws that must penetrate the intended substrate without disturbing the water-sediment boundary layer, close when positioned properly on the bottom, and retain a discrete sample of sediment while it is brought to the surface for processing. Hinged doors on the top of the grab prevent washout during sample retrieval. The grab is made of 12- to 20-gauge brass or stainless steel and weighs approximately 32-kg. The box-like part holding the sample has spring-operated jaws on the bottom that must be manually set. The sampler is available in several sizes; however, in very soft substrates only a tall model should be used, either a 23-cm or a 30.5-cm model. The Ekman grab can be operated from a boat with a winch and cable.

Smith-McIntyre Grab:

The Smith-McIntyre grab sampler is designed to obtain samples of macroinvertebrates from sediments in rough weather and deep water in estuaries and oceans. This device samples a surface area of 0.1-m^2 and is useful for sampling macroinvertebrates from a broad array of sand, gravel, mud, clay, and similar substrates.

The Smith-McIntyre grab sampler has hinged top doors to prevent sample washout and the pressure wave in descent. Its paired jaws are forced into the intended substrate by two "loaded" strong coiled springs when the grab

touches the bottom. The jaws close when positioned properly on the bottom, and retain a discrete sample of sediment to be brought to the surface for processing. The device is heavy and can weigh 45.4-kg or more. The chief advantage of the sampler is its stability and easier control in deep and rough waters. The spring-loaded jaws of the Smith-McIntyre grab must be considered a hazard and caution should be exercised when using the device. Due to the weight and size, this device must be used from a vessel with boom and lifting capabilities.

Modified Van Veen Grab:

The modified Van Veen grab sampler is used to obtain samples of macroinvertebrates from sediments in estuaries and other marine habitats. This device is useful for sampling sand, gravel, mud, clay and similar substrates and is available in three sizes: 0.06-m^2, 0.1-m^2, and 0.2-m^2. Larger versions of this grab are available, and their use is dependent upon the type of bottom to be sampled, and the type of vessel available to deploy the sampler.

The modified Van Veen grab sampler has paired jaws that penetrate the intended substrate without disturbing the water-sediment boundary layer. They are closed by the pincher-like action of two long arms. The long arms give added leverage for penetrating hard sediments.

The modified Van Veen is basically an improved version of the Petersen grab in that long arms have been attached to the jaws to help stabilize the grab on the bottom in the open sea just prior to or during closure of the device. This grab is used extensively in Puget Sound for the ambient monitoring program and for pollution-related surveys. Large hinged screen doors with rubber flaps have

been added to the top of the sampler for access to the surface of the sample. Additional weights can be applied to the modified Van Veen jaws to effect greater penetration in sediments, although penetration is not as deep in hard sand or cobble as with the Young grab or the Smith-McIntyre.

Young Grab:

The Young grab sampler is similar in operation to the Van Veen and the Smith-McIntyre, but the sample can be accessed undisturbed from the top of the grab through hinged doors like a Smith-McIntyre. It is encircled by a ring-like frame which enhances flat, stable landings of the grab on the substrate. Weights can be added to the frame to aid penetration in hard sand or cobble. A major advantage of the Young grab is efficient performance without the risk of injury associated with the spring-loaded Smith-McIntyre. This grab can be provided in a 0.044-m^2 and a 0.1-m^2 version. The former is appropriate to small boat operations while the latter size is more effective for marine work and obviously requires fewer lowerings or "drops" to obtain the same volume of material and community representation.

Recent comparisons of the Young and Smith-McIntyre grabs in rough Atlantic waters revealed consistently greater volumes of sediment collected by the Young grab in six trials each in soft sandy muds, sand, packed sand, and sand and gravel sediments. While the grabs were the same size (0.1-m^2) and had the same weight attached, the significant factor in performance was the design differences of the two grabs (Gibson 1995, unpublished).

While either the 0.1-m^2 Young or Smith-McIntyre designs are effective off-shore grabs for the biocriteria development purposes of this guidance, the Smith-

McIntyre provides better access to the sample while the Young grab is easier and safer to operate, especially in rough weather. An advantage of both designs is that the retrieved sample can be cross-sectioned and examined intact, although this is easier with the Smith-McIntyre design.

Core Samplers

Core samplers use a surrounding frame to ensure vertical entry; vertical sectioning of the sample is possible (USEPA 1986-1991). Coring devices can be used at various depths in any substrate that is sufficiently compacted so that an undisturbed sample is retained; however, they are best suited for sampling the relatively homogenous soft sediments, such as clay, silt, or sand of the deeper portions of estuaries and coastal marine waters. Because of the small area sampled, data from coring devices are likely to provide very imprecise estimates of the standing crop of macrobenthos.

KB, Ballcheck, and Phleger Corers:

KB type, Ballcheck, and Phleger corers are examples of devices used in shallow or deep water; they depend on gravity to drive them into the sediment. The cores are designed so that they retain the sample as it is withdrawn from the sediment and returned to the surface. Hand corers designed for manual operation are used in shallow water. Sections of the core can be extruded and preserved separately or the entire core can be retained in the tube and processed in the field or laboratory. Intact cores can also be preserved by freezing and processed later.

Additional replication with corers is feasible because of the small amount of material per sample that must be handled in the laboratory. Multiple-

head corers have been used in an attempt to reduce the field sampling effort that must be expended to collect large series of core samples (Flannagan 1970).

The Dendy inverting sampler (Welch 1948) is a highly efficient coring-type device used for sampling at depths to 2- or 3-m in nonvegetated substrates ranging from soft mud through coarse sand. Because of the small surface area sampled, data obtained by this sampler suffer from the same lack of precision (Kajak 1963) as the coring devices described above. Since the per-sample processing time is reduced, as with the corers, large numbers of replicates can be collected.

Stovepipe-type devices include the Wilding sampler (Wilding 1940, APHA 1992) and any tubular material such as 60- to 75-cm sections of standard 17-cm diameter stovepipe (Kajak 1963) or 75-cm sections of 30-cm diameter aluminum irrigation pipe fitted with handles. In use, the irrigation pipe or commercial stovepipe is manually forced into the substrate, after which the contained vegetation and coarse substrate materials are removed by hand. The remaining materials are repeatedly stirred into suspension, removed with a long-handled dipper and poured through a wooden-framed floating sieve. Because of the laborious and repetitive process of stirring, dipping, and sieving large volumes of material, the collection of a sample often requires 20- to 30-minutes.

The use of stovepipe samplers is limited to standing or slowly moving waters having a maximum depth of less than 60-cm. Since problems relating to depth of sediment penetration, changes in cross-sectional area with depth of penetration, and escape of organisms are circumvented by stovepipe samplers,

they are appropriate for quantitative sampling in all shallow-water benthic habitats and can be deployed from small boats. They probably represent the only quantitative device suitable for sampling shallow-water habitats containing stands of rooted vascular plants and they will collect organisms inhabiting the vegetative substrates as well as those living in sediments.

In marine waters, benthic macrofauna are generally collected using various box cores deployed from ships or other platforms, or diver operated cores. A box coring device consisting of a rectangular corer having a cutting arm which can seal the sample prior to retraction from the bottom should be used. In order to sample a sufficient number of individuals and species, and to integrate the patchy distribution of fauna, each sample should have a surface area of no less than 100-cm^2 and a sediment depth of at least 20-cm. In sediments having deep, burrowing fauna, a box corer capable of sampling deeper sediment may be needed. In sandier sediments, it may be necessary to substitute a grab sampler for the box corer in order to achieve adequate sediment penetration. Visual inspection of each sample is necessary to insure that an undisturbed and adequate amount of sample is collected.

Sieve Mesh Size

The use of different sieve mesh sizes for screening benthic samples limits the comparability of results between marine monitoring studies (Reish 1959; Rees 1984). The major advantage of using a smaller mesh size is the retention of both juvenile and adult organisms as well as large-and small-bodied taxa. The major disadvantage is the concomitant increased cost of sample processing. For example, using a 0.5-mm mesh rather than a 1.0-mm mesh could increase

retention of total macrofaunal organisms by 130 to 180%; however, costs for processing the samples may increase as much as 200% (USEPA 1986-1991).

It is recommended that a standard mesh size be selected for all monitoring studies. A review of estuarine monitoring programs from around the country (Bowman et al. 1993) showed that both 0.5- and 1.0-mm mesh sizes are used, with a slight majority of the programs reviewed using a 0.5-mm mesh screen (Table 5-5). Dauer (1993) evaluated biocriteria developed from data collected as part of the Virginia Benthic Biological Monitoring Program using a 0.5-mm mesh screen.

Sieving can be done either aboard the survey vessel or on shore after the cruise. Sieving occurs prior to fixation (sample preservation) aboard the vessel, whereas waiting until after the cruise requires fixation prior to sieving. If inadequate concentrations of fixative are added and deterioration or decomposition of organisms occurs, there may be significant sample degradation. If large numbers of samples are to be collected, field sieving reduces sample storage requirements as well as the modification/loss of data (USEPA 1992, 1994d).

After samples have been collected, the samples must be processed so that data can be collected and analyzed. Two aspects of sample processing of particular concern are the subsampling and identification that may occur in the field or laboratory. Sorting procedures are described in Klemm et al. (1992).

▸ Subsampling of benthic infauna can be accomplished by subcoring; i.e., removing smaller core samples from within a grab or core sample, and sorting all organisms found within the subcore. The size and number of subcores that should be taken will depend upon the variability of the infaunal community. Representative subsampling can be difficult to achieve if benthic species have patchy or clumped distributions. Subsampling can also damage collected organisms (e.g., polychaete worms), decreasing the number of specimens that can be identified to genus or species;

▸ Several studies have examined the effect of varying levels of taxonomic analysis on the results of statistical measures of the infaunal community (e.g., Ferraro and Cole 1990, 1992, 1995, Warwick 1988, Warwick et al. 1990). The studies indicate that in some instances species-level taxonomic identification does not yield any more information than family- or even phylum-level identification. The degree of taxonomic proficiency required to adequately characterize the community will depend upon the diversity present in the community. Species level identification is necessary and cost-effective for fish surveys. However, while this is desired for macroinvertebrates, it is often too costly and assessment needs can usually be met at the genus level.

Although species-level identifications may not be necessary for classifying sites as minimally impaired or impaired, this degree of taxonomic identification may be required to assess the sources of impairment using data collected in Tier 3. Species-level identifications require greater taxonomic expertise than do higher taxonomic divisions; this species level of expertise may not be as readily available to state agencies. If this is the case, then state resource managers must determine whether the cost of

Table 5-5.　　　　Mesh sizes used in estuary benthic monitoring programs.

Monitoring Program	Mesh Size (mm)	Reference
Chesapeake Bay	0.5	Dauer 1993, Holland et al. 1987, 1988, 1989, Ranasinghe et al. 1992
Tar/Pamlico	0.5	Eaton 1992a-d
EMAP-Near Coastal	0.5	USEPA 1992, Weisberg et al. 1993, Holland 1990
Naples Bay, Florida	0.595	Simpson et al. 1979
Puget Sound Ambient Monitoring Program	1.0	PSWQA 1988, 1990, 1991
Puget Sound Estuary Program	1.0	Simenstad et al. 1991

contracting these identifications is justified based on the information obtained and the assessment tier to which it would be applied. One approach to this problem of obtaining sufficient taxonomic expertise is for the states of a region to cooperate in a joint venture to employ the taxonomic expertise necessary to all. In this manner the cost of a skilled taxonomist, either contracted or on staff, can be shared.

5.1.2　Fish

Fish communities include species in a variety of trophic levels (omnivores, herbivores, planktivores, piscivores). Fish are long-lived, integrate long- and short-term changes, and they also integrate effects of lower trophic levels; thus, fish community structure is a good measure of integrated environmental health. Estuarine and coastal marine fish receive a large amount of public attention because of sport and commercial fishing and attendant concerns regarding fish production and safety for human consumption. On the negative side, fish may be wide-ranging or migratory and might not reflect local conditions in estuaries and coastal marine waters; some fish species may also be influenced by management (stocking), angling, and commercial harvesting; and unbiased sampling is difficult because each feasible gear type is highly selective.

Sampling Gear

Fish communities may vary considerably among the numerous habitat types that may be present in a target estuary or coastal marine area. The choice of sampling method and gear type will depend upon the habitat and the fish species of interest. Shallow areas may best be sampled using dip nets or beach seines, while deeper waters may be sampled using gill nets, purse seines, or otter trawls. Net and mesh size should be appropriate to allow a representative sample of target fish to be obtained. Fishing effort should be comparable among stations with constant tow distances, times, speeds, and lengths of trawl warps. Because there is no easy way of estimating population size in any given area of an estuary or coastal marine area, consistency in effort is of the utmost

importance to allow legitimate comparisons among sites.

Maryland DNR's IBI sampling techniques are designed to sample the nearshore fish communities in the tidal tributaries of the Chesapeake Bay. They were modeled after the Maryland Striped Bass Juvenile Seine Survey which has been ongoing since 1954 (Goodyear 1985). Two beach seines are pulled at each site allowing a half hour interval between hauls for repopulation of the seine area. Seines are pulled with the tide employing a "quarter sweep" method where one end of the seine is held on shore while the other end is fully extended perpendicular to shore, and then pulled back into shore forming a semi-circle. The seine used is a bagless 6.4-mm mesh seine 30.5-m in length and 1.2-m deep. Precautions are taken upon approaching the site to avoid disturbance of the sampling area.

Concurrent trawls are pulled with the tide in the channel adjacent to shore. A small otter trawl (3.1-m with 12.8-mm stretch mesh, and 50.8-cm x 25.4-cm doors) with tickler chains is used to sample the bottom community local to the seine sample area. Water quality measurements (temperature, dissolved oxygen, pH, conductivity, and salinity)

and Secchi depth are also taken in the trawling area. Water quality is sampled at surface, mid, and bottom depths. These measurements have proven useful in relating water quality parameters to fish communities. A summary of fish sampling is given in Table 5-6.

▸ Subsampling of fish collected using any of the sampling methods mentioned above is problematic. It is probably most efficient and statistically valid to identify and make external measurements and observations of all fish caught during a given tow or time period.

▸ The level of taxonomic identification required to effectively characterize the fish community will depend upon the diversity of the community being sampled and the metrics being used to evaluate the data. Identification to species is preferable for most individuals taken in a given area. Individuals that cannot be field-identified should be preserved and returned to the lab for identification.

5.1.3   Aquatic Macrophytes

Macrophytes form an integral part of the littoral zone of many estuaries and

Table 5-6.   Sampling summary for fish.

Habitat	Sublittoral.
Sampling Gear	Seines and any gear that effectively captures bottom-feeding and pelagic fish, usually otter trawls.
Index Period	Any season can be selected depending upon migration and recruitment patterns in the region. Seasonal sampling might be needed to assess particular problems.
Sampling	Bottom-feeding and pelagic fish. Sufficient sets of gear to obtain representative species counts (usually 4 or more).
Analysis	Collected species are weighed, measured, and examined for external abnormalities (lesions, growths, deformities). Histopathology may be performed.

coastal marine waters, serving as habitat for fish and invertebrates as well as being a distinct biological assemblage. For many estuaries, the areal extent and distribution of SAV is used as an indicator of estuarine quality (Batiuk et al. 1992). Ecosystems whose primary producer component is dominated by aquatic macrophytes can be transformed to macro algae or phytoplankton-dominated systems through nutrient enrichment. Increased nutrient input stimulates macrophyte growth; however, it also promotes growth of periphyton and phytoplankton, which shade the SAV. The shading reduces macrophyte growth and survival (Dennison et al. 1993, Batiuk et al. 1992). Overall, macrophyte standing stock is an excellent indicator of estuarine water quality. The presence of confounding factors, such as diseases, can be determined from examination of affected plants, or from historical information. Potential macrophyte metrics are listed in Table 5-7 and the recommended sampling protocol for macrophytes is summarized in Table 5-8. Field sampling can be performed in a single visit. Plants are identified and weighed on-site, with voucher specimens preserved as necessary. There is no intensive laboratory analysis required.

5.1.4 Phytoplankton

Phytoplankton are the base of most estuarine food webs (Day et al. 1989), and fish production is linked to phytoplankton primary production (e.g., Day et al. 1989). Excessive nutrient and organic inputs from human activities in estuaries and their watersheds leads to eutrophication characterized by: reduction in seagrasses, increases in phytoplankton biomass, macrophyte biomass (macroalgal biomass), reduced water clarity, and reduced oxygen saturation in bottom waters. From a human perspective, problems might include loss of aesthetic appeal, decreases in desirable commercial and game fishes, and loss of recreational access caused by increased macrophyte production.

Phytoplankton standing stock is measured by surface chlorophyll *a* concentration, sampled at the 0.5-m depth at each sampling site (Table 5-9). Tiers 1 and 2 can use a single measurement taken at each sampling site with a fluorometer attached to a conductivity-temperature-depth meter (CTD) (USEPA 1994c) taken from June through September. Alternatively, chlorophyll *a* may be determined spectrophotometrically on phytoplankton samples returned to the lab. Tier 2 can include identification of dominant taxa, including nuisance taxa. Tier 3 uses a seasonal or annual average surface chlorophyll concentration from all stations over all sampling events and can include full characterization of the phytoplankton community.

If phytoplankton communities are to be sampled, several techniques may be employed; these are described more fully in APHA (1992).

▸ Phytoplankton samples may be obtained using water bottles deployed on a wire at a given, or preferably various, depths. The water bottles used should be constructed and cleaned in a manner appropriate for the collection of phytoplankton samples (e.g., Niskin bottles washed and rinsed in order to remove contaminants). Chlorophyll concentration is measured from the sampled water, and phtyoplankton cells may be filtered or settled for identification and enumeration.

Table 5-7.   Potential aquatic macrophyte metrics.

Metric	Response to impairment
Tier 1: % cover dominant taxa	substantially more or less than reference substantially more or less than reference
Tiers 2-3: % cover biomass maximum depth of plant growth density of new shoots stem counts	reduced or enhanced substantially more or less than reference reduced under enrichment reduced reduced

Table 5-8.   Sampling summary for aquatic macrophytes.

Habitat	Euphotic zone.
Sampling Gear	Aerial photography; quadrats
Index Period	During growing season
Sampling	Tier 1: Estimate of area covered by macrophytes. Tiers 2-3: Quadrat samples for biomass collected by diver; 3-5 randomly placed transects perpendicular to shore; samples are taken at 0.5-m depth intervals from edge of emergent zone to the sublittoral.
Analysis	Tier 1: Dominant taxa identified, % cover estimated from aerial photography. Tiers 2-3: All species identified, relative abundance of each estimated from wet weight.

Table 5-9.   Sampling summary for phytoplankton.

Habitat	Each sampling site preferred.
Sampling Gear	Fluorometer attached to CTD (USEPA 1994e) for *in situ* measurements; or spectrophotometrically on water samples collected with a water sampler.
Index Period	Tiers 1 and 2: June - September Tiers 2 (optional) and 3: growing season average; 6-10 samples; March - October (longer in subtropical regions).
Sampling	Preferred: single sample, 0.5-m depth. Alternate: at same depths as nutrient samples.
Analysis	Tier 1: Chlorophyll *a* mg/L (Tiers 1-3). Tier 2: ID dominant taxa. Tier 3: full community species characterization.

- Phytoplankton may also be collected by net hauls using a plankton net with an appropriate mesh size.

Bottle collections are most useful when analyzing a bulk community measure such as chlorophyll *a* concentration (assuming a fluorometer coupled to a CTD is not used), while net hauls are better for studies designed to enumerate species. Water samples for chlorophyll *a* determination can also be used for nutrient analysis.

- The level of taxonomic identification that should occur will depend upon the diversity of the community, the analyses that are to be performed, and the cost and availability of taxonomic experience;

- If phytoplankton are collected using water bottles, the water may be subsampled in the field or lab prior to analysis. The size and number of subsamples that should be taken will depend upon the variability present in the community;

- If subsamples are taken from net hauls, it may be necessary to resuspend the organisms found in the cod end of the net in a larger volume of water in order to facilitate subsampling.

5.1.5 Zooplankton (Developmental)

Zooplankton are most effectively sampled using net hauls with 118-μm mesh sizes. Because zooplankton are known to exhibit diel periodicity in their locations in the water column, sampling times should reflect this temporal variability; i.e., sampling should, in general, be conducted at night. Also, consideration should be given to the use of vertical or oblique tows. In any instance, gear size, mesh size, rate of retrieval on the haul back, vertical or oblique tow, time of day or night and tide cycle are factors which must be kept constant if zooplankton surveys are to be included in biocriteria development.

Meaningful bulk community measurements do not exist for zooplankton; therefore, if zooplankton are to be sampled, they should be identified and enumerated. It may be difficult to locate or develop the taxonomic expertise necessary to identify zooplankton to species, especially given the large number of planktonic larvae. Zooplankton are considered to be in a developmental status with respect to their use as an estuarine and coastal marine bioassessment assemblage. Zooplankton populations experience year-round seasonal fluctuations in abundance as a result of variable larval recruitment into the population, variable food sources, and physical processes which may move larvae and adults into and out of the estuary (Day et al. 1989). The pattern of seasonal abundance differs with changes in latitude. Zooplankton in higher latitudes have one or more mid-summer peaks and very low numbers during the winter.

Abundances in temperate estuaries are much more variable and may experience spring peaks and minima during the summer and winter months. Tropical estuaries do not experience the low in population during the winter.

Some long-term monitoring projects have identified community measures that indicate changes in

environmental conditions over time (e.g., nutrient loads or toxicants), as well as particular zooplankton taxa whose densities affect larval fish survival (Buchanan 1991). Zooplankton community characteristics that are under investigation for application as bioindicators include:

▸ Diversity, measured through standard indexes such as Shannon-Wiener, to evaluate the taxonomic complexity of the assemblage;

▸ Ratios of specific taxonomic groups within the assemblage to gauge community balance and identify possible impairment;

▸ Presence of Hypotrichs (a ciliate of the order Hypotrichida);

▸ Total biomass to assess assemblage production;

▸ Relative abundance of pollution tolerant and sensitive species to identify and evaluate impairments to the assemblage;

▸ Unnatural variability in abundance can be used to identify the presence of short-term pollution or climate events;

▸ Size structure can be used to evaluate the growth of cohorts in the assemblage, which can provide information on possible short- and long-term system perturbations.

5.1.6 Epibenthos (Developmental)

The epibenthos assemblage is also considered to be in a developmental stage for use in estuarine and coastal marine bioassessment. Taxa within

the epibenthic community appear to be persistent and sensitive to environmental stress. They are characterized by physiological mechanisms that allow them to tolerate the varying salinity, DO, and temperature conditions encountered in estuaries and coastal marine waters, or reproductive cycles that allow them to avoid high-stress periods. Some epibenthos and facultative infauna can relocate to avoid areas of environmental stress.

Epibenthos can be sampled using a Renfro beam trawl, otter trawl, or epidbenthic sled. Camera tows or remotely operated vehicles with camera or video capabilities may also allow enumeration of epibenthos, although collection of organisms would not be possible and quantitative assessments difficult. Subsampling might involve a process similar to that suggested by Plafkin et al. (1989); a box with a numbered grid system into which collected epibenthos are evenly distributed could be used to randomly select an appropriate number of organisms for subsequent sorting.

Some of the advantages to using epibenthos for estuarine and coastal marine bioassessment are:

▸ This assemblage is very sensitive to anthropogenic sources of stress, and it can be used in both a nearfield and farfield context with equal facility;

Sampling can be conducted in shallow waters using a dip net and in deep waters with a trawl;

▸ The total number of common species will be limited by the fact that the deep water sampling gear is restricted to fairly level bottoms;

▸ Subsampling can be employed to reduce labor costs and increase cost-effectiveness;

▸ Field and lab work, and data analyses can be done quickly with trained personnel;

▸ Samples can be sorted qualitatively, and a nonparametric analysis can be applied to provide a quick screening method.

▸ Seagrasses and macroalgae can hinder or increase the time necessary for field sorting;

▸ The seasonality of epifauna needs to be factored into the sampling design.

The developmental method described in Chapter 13 appears promising for detecting impairment. If successfully adapted to regions outside Florida, North Carolina, and Puget Sound where it is being tested, it may become a standard estuarine bioassessment method in the future. A proposed sampling protocol is summarized in Table 5-10.

Table 5-10. Sampling summary for epibenthos.

Habitat	Soft sediments (sleds and trawls); shallow, vegetated (dip net)
Sampling Gear	Renfro Beam Trawl (Farrell 1993a,b), small otter trawl; epibenthic sled; dip net
Index Period	Preferred: mid-summer Alternative: growing season, average of 10 samples.
Sampling	Ca. 4-m tow length in estuaries; 0.1 - 0.5 nm tow lengths (DGPS) in coastal waters and Puget Sound.
Analysis	Taxonomic ID preferably to species.

The disadvantages of this assessment methodology are:

▸ The stress index is developed solely for anoxia; it might not allow assessment of other stressors;

▸ Stress values may not be available for many species, or may be difficult to determine;

▸ Sleds and trawls are restricted to level bottoms; and cannot be used for sampling hard bottoms, or rock rubble;

5.1.7 Paleoenvironmental Systems (developmental)

Diatom and foraminifera species have narrow optima and tolerances for many environmental variables, which make them useful in quantifying environmental characteristics to a high degree of certainty. They immigrate and replicate rapidly, which makes them quick to respond to environmental change (Dixit et al. 1992). Changes in assemblages also correspond closely to shifts in other biotic communities sampled in estuaries such as aquatic macrophytes,

zooplankton, and fish. They have also been used alone as environmental indicators of eutrophication, metal contamination, salinification, thermal effluents, and land use changes. Furthermore, since diatoms and foraminifera are abundant in almost every marine ecosystem, a relatively small sample is sufficient for analysis. This allows for many samples to be easily collected, analyzed, and archived (Dixit et al. 1992).

The general lack of time-series data has prompted attempts to demonstrate marine eutrophication from present-day observations using the benthic community and chemical criteria (Dale et al. 1999). Benthic foraminifera have been proven useful as indicators of oxygen concentration in bottom sediments (Alve 1991). Dinoflagellate cysts are also increasingly useful as indicators of short-term environmental change caused by climate and human pollution (Dale et al. 1999). The cysts are recovered by pollen identification techniques; they are acid-resistant and therefore not subject to dissolution problems sometimes affecting diatoms and foraminifera (Dale et al. 1999). Measurements of biogenic silica in sediments are most often used as an index of diatom production (Stoermer et al. 1990, Conley et al. 1993, Cooper 1995). Isolation of BSi from Si in mineral phase is based upon the fact that the silica of diatoms is only weakly crystalline and dissolves readily in a weak base. Potential indicators and a proposed sampling summary are shown in Tables 5-11 and 5-12.

The total number of cores taken in a particular estuary is dependent upon the hydrological complexity of the estuary. Generally, one to three cores, but some times up to ten are required

from each estuary or tributary being assessed. However, once a paleoecological record is established, there is no need to repeat the sampling.

Although the number of cores is small, each core requires substantial effort to analyze: sectioning, radioisotope dating, chemical analysis, pollen analysis for further dating, and diatom or foraminifera analysis. Current estimates for paleological analysis is about $100 per section (not per core), depending on the number and intensity of analysis done on each section and the experience of the lab performing the analysis. The complexity of estuaries requires some background information about the area in which sampling is occurring. This information should assist in decision making on the location and number of cores to be retrieved.

The study of paleoenvironmental systems requires a corer that will retrieve an intact core, with minimal edge disturbance (Table 5-4). K-B, Phleger, and Piston corers have all been used successfully for these analyses (see Section 5.1.1). Small surface area is not an issue; a single core will suffice.

5.2 Sampling Design Issues

Consideration of sampling design is critical in developing a new monitoring program for estuarine bioassessment and biocriteria. Sampling design includes defining the questions to be addressed by the data, defining the units that will be

Table 5-11. Potential paleoecological indicators

Indicator	Response to Impairment	Reference
Taxa richness (diatom, foraminifera, dinoflagellate cysts)	reduced	Cooper and Brush 1991
Biogenic silica	increase with nutrient enrichment	Turner and Rabalais 1994
Total organic carbon, Total N, Total S	increase with enrichment	Turner and Rabalais 1994
Ammonia/Elphidium ratio (foraminifera)	increase with hypoxia	Sen Gupta et al. 1996
Centric/pennate ratio (diatoms)	increase with nutrient enrichment	Cooper and Brush 1991
% *Cyclotella*	increase with nutrient enrichment	Cooper and Brush 1991
sedimentation rate	increase with watershed erosion	Brush 1989
Dinoflagellate cysts	increase with cultural eutrophication	Dale et al. 1999
% *Fursenkoina*	increases with hypoxia	Alve 1991
% *Trochammina*	increases with hypoxia	Patterson 1990

Table 5-12. Sampling summary for paleoenvironmental systems

Habitat	Stable depositional zone, biogeochemical conditions for preservation
Sampling gear	Bottom corer
Index period	None
Sampling	Tiers 1-2: none
	Tier 3: background information specific to the estuary being sampled will determine the number of cores necessary.
Analysis	Cores sectioned at regular intervals depending on deposition rate and resolution desired.
Diatoms Foraminifera Dinoflagellate Cysts	Species composition and enumeration of at least 300 organisms in each section. Digestion/clarification methods depend on assemblage.
Age of sections up to 150 years	^{210}Pb determination based on radioisotope assay with alpha spectroscopy.
Older than 150 years	Palynological (pollen) analysis correlated with known historical changes in terrestrial vegetation (land use), and ^{14}C analysis (>1000 yr).

Sampling Program Issues

sampled, and developing a sampling design that is cost-effective for answering the defined questions.

5.2.1   Statement of the Problem

The first task in developing a sampling and assessment program is to determine, and be able to state in simple fashion, the principal questions that the sampling program will answer. Questions may or may not be framed as hypotheses to test, depending on program objectives. For example, suppose that a sampling program objective is to establish reference conditions for biological criteria for estuaries in a given region. Typically, the initial objectives of a survey designed to develop criteria are to identify and characterize classes of reference sites in estuaries. Initial questions may then include:

▸   Should minimally disturbed sites be divided into two or more classes that differ in biological characteristics and dynamics?;

▸   What are the physical, chemical, and relevant biotic characteristics of each of the estuary site classes?

After the monitoring and assessment program has developed biological criteria, new questions need to be developed that encompass assessments of individual sites, estuaries, or estuaries of an entire region or state. Specific questions may include:

▸   Is site abc similar to reference sites of its class (unimpaired), or is it different from reference sites (is it altered or impaired)?;

▸   Overall, what is the status of estuarine waters in the region? What percentage of estuarine

waters is similar to reference conditions? What percentage is impaired?;

▸   Has estuary abc changed over a certain period? Has it improved or deteriorated?;

▸   Overall, have estuarine waters in the region improved or deteriorated over a certain period? Have individual estuaries improved? Are more waters similar to reference conditions now than some time ago?

Finally, resource managers often wish to determine the relationships among variables, that is, to develop predictive, empirical (statistical) models that can be used to design management responses to perceived problems. Examples of specific questions include:

▸   Can trophic state of an estuary be predicted by areal nitrogen loading rate?;

▸   Can biota of an estuary be predicted by watershed land use?

Monitoring and assessment data, and derived models, may also be used to help determine causal relationships between stressors and responses of systems. Inferring cause requires manipulative experiments, or inference from multiple lines of evidence (Suter 1993). Since surveys and monitoring programs preclude experimental investigations, inference of causal relations is beyond the scope of this document. Often, there is enough experimental evidence available from other studies so that additional causal experiments are not necessary and would be superfluous (e.g., current knowledge of nutrients and trophic state generally makes it

---

unnecessary to "prove" experimentally which nutrients are limiting). The development of predictive models usually does not require formal hypothesis testing.

It is also necessary to specify the units for which results will be reported. Usually, these units are the population (e.g., all estuarine waters), but often subpopulations (e.g., embayments or tributaries of a given estuary) and even individual locations (e.g., sites of special interest) can be used. In order to help develop the sampling plan, it is useful to create hypothetical statements of results in the way that they will be reported, for example:

▸ Status of a place: *Baltimore harbor is degraded;*

▸ Status of a region: *20% of the area of Puget Sound has elevated trophic state, above reference expectations;* or *20% of estuaries in Oregon have elevated trophic state;*

▸ Trends at a place: *Benthic species richness in Baltimore harbor has increased by 20% since 1980;*

▸ Trends of a region: *Average estuary trophic state in New Jersey has increased by 20% since 1980;* or *Average benthic index values in 20% of estuaries of the west coast have increased by 15% or more since 1980;*

▸ Relationships among variables: *50% increase of N loading above natural background is associated with decline in taxa richness of benthic macroinvertebrates, below reference expectations;* or *Estuaries receiving runoff from large urban areas have 50% greater probability of elevated trophic state above reference than estuaries not receiving such runoff.*

5.2.2 Definition of the Assessment Unit

Defining the resource and assessment unit of the resource begins the process of developing biological criteria. An "assessment unit" is a whole estuary or part of an estuary, that will be assessed as meeting criteria, being impaired, etc. Clearly, a single square meter where a grab sample is taken is not large enough to be an assessment unit. An assessment unit should consist of a definable segment, basin, or entire estuary. For example, a large complex estuary such as Puget Sound could be divided into its component inlet bays, canals, and passes. Many of the larger components could in turn be divided into segments.

Segmentation could be determined by some combination of mean salinity, water residence time, dominant substrate, or mean depth. For example, since estuarine fauna are determined by salinity, segmentation often corresponds to salinity zone (tidal fresh, oligohaline, mesohaline, polyhaline, and marine). Small estuaries, such as salt ponds in New England, could be single assessment units.

An assessment unit is the smallest spatial subdivision of an estuary that will be assessed; i.e., given a rating of good or poor. An assessment may be based on one or more sample units within an assessment unit. A sample unit (or sample site) is a site where an observation is made.

5.2.3 Specifying the Population and Sample Unit

Sampling is statistically expressed as a sample from a population of objects. Thompson (1992) suggested in some cases, the population is finite,

countable, and easy to specify, (e.g., all persons in a city, where each person is a single member of the population). In estuaries, the population is often more difficult to specify and may be infinite, (e.g., the sediment of San Francisco Bay, where any location in the Bay defines a potential member of the population). Sampling units may be natural units (entire estuaries, cobbles on a beach), or they may be arbitrary (plot, quadrat, sampling gear area or volume) (Pielou 1977). Finite populations may be sampled with corresponding natural sample units, but often the sample unit (say, an estuary) is too large to measure in its entirety, and it must be characterized with one or more second stage samples of the sampling gear (bottles, benthic grabs, quadrats, etc.).

The objective of sampling is to best characterize individual sample units in order to estimate some attributes (e.g., nutrient concentrations, DO) and their statistical parameters (e.g., mean, median, variance, percentiles) of a population of sample units. The objective of the analysis is to be able to say something (estimate) about the population. Examples of sample units include:

▸ A point in an estuary (may be characterized by single or multiple sample device deployments). The population would then be all points in the estuary, an infinite population. This is the most common sample unit applied to estuarine assessments;

▸ A constant area, (e.g., square meter, hectare). The population would be an artificial one consisting of all square meters of estuarine surface area in an estuary, a state or a region;

▸ An estuary or a definable portion of the estuary as a single sample unit. Whole estuaries as sample units would only be used in very broad-scale regional assessments, as was done by EMAP-NC, for example, for small estuaries as a population (e.g., Strobel et al. 1995).

5.2.4   Sources of Variability

Variability of measurements has many possible sources, and the intent of many sampling designs is to minimize the variability due to uncontrolled or random effects, and conversely to be able to characterize the variability caused by experimental or class effects. For example, we may stratify estuarine waters by salinity and bottom substrate type (rocky, sandy, muddy). Typically, we stratify so that observations (sample units) from the same stratum will be more similar to each other than to sample units in other strata.

Environmental measures vary across different scales of space and time, and sampling design must consider the scales of variation. When sampling estuaries, measurements (say, benthic assemblages) are taken at single points in space and time (1 point along a transect in mid-summer). If the same measurement is made at a different place (littoral zone), embayment, or time (winter), the measured values will likely be different. A third component of variability is the ability to accurately measure the quantity interested in, which can be affected by sampling gear, instrumentation, errors in proper adherence to field and laboratory protocols, and the choice of methods used in making determinations.

The basic rule of efficient sampling and measurement is to sample so as to minimize measurement errors; to maximize the components of variability that have influence on the central questions and reporting units; and to control other sources of variability that are not of interest, that is, to minimize their effects on the observations. Many locations are sampled in order to examine and characterize the variability due to different locations (the sampling unit). Each site is sampled in the same way, in the same place, and in the same time frame to minimize confounding variability.

In statistical terminology, there is a distinction between sampling error and measurement error that has little to do with actual errors in measurement. Sampling error is the error attributable to selecting a certain sample unit (e.g., an estuary or a location within an estuary) that may not be representative of the population of sample units. Statistical measurement error is the ability of the investigator to accurately characterize the sampling unit. Thus, measurement error includes components of natural spatial and temporal variability within the sample unit as well as actual errors of omission or commission by the investigator. Measurement error is minimized with methodological standardization: selection of cost-effective, low variability sampling methods, proper training of personnel, and quality assurance procedures to minimize methodological errors. In analytical laboratory procedures, measurement error is estimated by duplicate determinations on some subset of samples (but not necessarily all). Similarly, in field investigations, some subset of sample units should be

measured more than once to estimate measurement error.

If the variance of individual measurements (measurement error) is unacceptably large; i.e., as large or larger than variance expected among sample units, then it is often necessary to alter the sampling protocol, usually by increasing sampling effort in some way, to further reduce the measurement error. Measurement error can be reduced by multiple observations at each sample unit, (e.g., multiple dredge casts at each sampling event, multiple observations in time during a growing season or index period, depth-integrated samples, or spatially integrated samples.

A less costly alternative to multiple measures in space is to make spatially composite determinations. In nutrient or chlorophyll determinations, a water column pumped sample, where the pump hose is lowered through the water column, is an example of a spatially composite determination. Spatial integration of an observation and compositing the material into a single sample is almost always more cost-effective than retaining separate, multiple observations. This is especially so for relatively costly laboratory analyses such as organic contaminants and benthic macroinvertebrates. Many estuarine programs have adopted sampling protocols consisting of multiple grabs at a site that are then composited into a single bucket for laboratory determinations (e.g., EMAP Near Coastal: 3 composited Van Veen grabs at each site; Holland 1990).

Statistical power is the ability of a given hypothesis test to detect an effect that actually exists, and must be considered when designing a

sampling program (e.g., Peterman 1990, Fairweather 1991). The power of a test (1-) is defined as the probability of correctly rejecting the null hypothesis (H_0) when H_0 is false; i.e. the probability of correctly finding a difference [impairment] when one exists. For a fixed confidence level (e.g., 90%), power can be increased by increasing the sample size or the number of replicates. To evaluate power and determine sampling effort, an ecologically meaningful amount of change in a variable must be set. See Chapter 12 for a discussion of statistical power, and examples.

Optimizing sampling design requires consideration of tradeoffs among the measures used, the effect size that is considered meaningful, desired power, desired confidence, and resources available for the sampling program. Every study requires some level of repeated measurement of sampling units to estimate precision and measurement error. Repeated measurement at 10% or more of sites is common among many monitoring programs.

5.2.5 Alternative Sampling Designs

Sampling design is the selection of a part of a population to observe the attributes of interest, in order to estimate the values of those attributes for the whole population. Classical sampling design makes assumptions about the variables of interest, in particular, it assumes that the values are fixed (but unknown) for each member of the population, until that member is observed (Thompson 1992). This assumption is perfectly reasonable for some variables, say, length, weight, and sex of members of an animal population, but it seems less reasonable for more dynamic

variables such as nutrient concentrations, loadings, or chlorophyll concentrations of estuaries. Designs that assume that the observed variables are themselves random variables are model-based designs, where prior knowledge or assumptions (a model) are used to select sample units.

Probability-based designs (random sampling)

The most basic probability-based design is simple random sampling, where all possible sample units in the population have the same probability of being selected, that is, all possible combinations of n sample units have equal probability of selection from among the N units in the population. If the population N is finite and not excessively large, a list can be made of the N units, and a sample of n units is randomly selected from the list. This is termed *list frame sampling*. If the population is very large or infinite (such as locations in an estuary), one can select a set of n random (x,y) coordinates for the sample.

All sample combinations are equally likely in simple random sampling, thus there is no assurance that the sample actually selected will be representative of the population. Other unbiased sampling designs that attempt to acquire a more representative sample include stratified, systematic, multistage, and adaptive designs (Figure 5-2). In stratified sampling, the population is subdivided or partitioned into strata, and each stratum is sampled separately. Partitioning is typically done so as to make each stratum more homogeneous than the overall population. Systematic sampling is the systematic selection of every k^{th} unit of the population from one or

Sampling Methods

Simple Random: Samples are independently located at random

Systematic: Samples are located at regular intervals

Stratified: The study area is divided into nonoverlapping strata and samples are obtained from each

Multistage: Large primary units are selected which are then subsampled
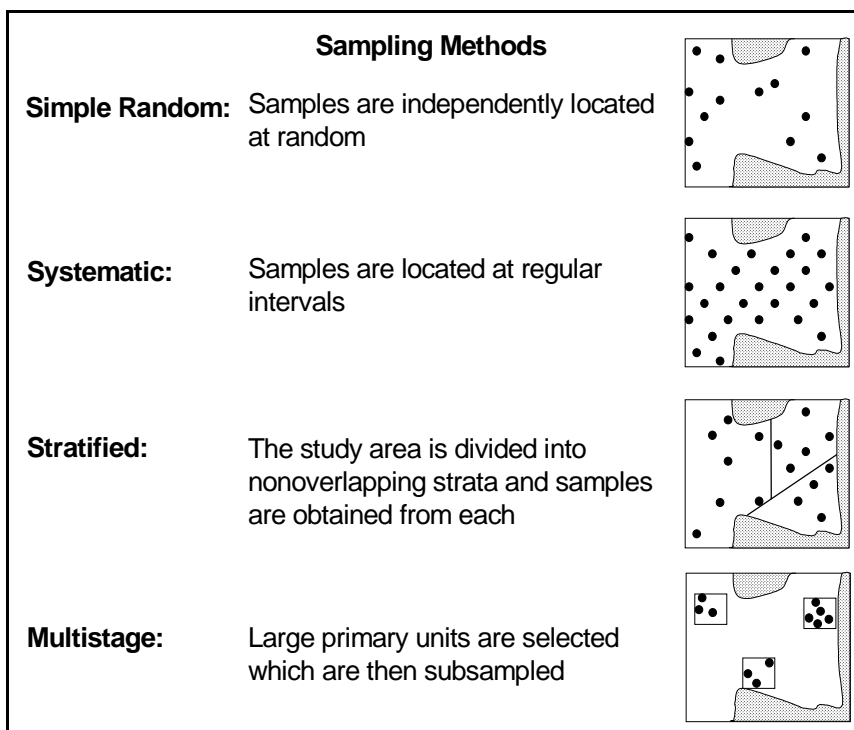
Figure 5-2

Description of various sampling methods. Adapted from USEPA 1992.

more randomly selected starting units, and ensures that samples are not clumped in one region of the sample space. Multistage sampling requires selection of a sample of large primary units, such as fields, hydrologic units, rectangles, or hexagons, and then selection of secondary sample units such as plots or estuaries within each primary unit in the first stage sample.

Estimation of statistical parameters requires weighting of the data with inclusion probabilities (the probability that a given unit of the population will be in the sample) specified in the sampling design. In simple random sampling, inclusion probabilities are by definition equal, and no corrections are necessary. Stratified sampling requires weighting by the inclusion probabilities of each stratum. Unbiased estimators have been developed for specific sampling designs, and can be found in sampling textbooks, such as Thompson (1992).

Model-based designs

Use of probability-based sampling designs may miss relationships among variables (models), especially if there is a regression-type relationship between an explanatory and a response variable. As an example, estimation of benthic response to discharge or outfalls requires a range of sites from those directly adjacent to the outfalls to those distant from, and presumably unaffected by, the outfalls (e.g. Warwick and Clarke 1991). A simple random sample of estuarine sites is not likely to capture the entire range, because there would be a large cluster of far sites, with few at high ends of the gradient. A simple random sample may therefore be highly inefficient with respect to models or specific hypotheses.

In model-based designs, sites are selected based on prior knowledge of auxiliary variables, such as estimated loading, depth, salinity, substrate

type, etc. These designs preclude an unbiased estimate of the state of the estuaries, unless the model can be demonstrated to be robust and predictive, in which case the population value is predicted from the model and from prior knowledge of the auxiliary (predictive) variables. Selection of unimpacted reference sites is an example of a model-based design which cannot later be used for unbiased estimation of the biological status of estuaries. Ideally, it may be possible to specify a design that allows unbiased estimation of both population and model, with an appropriately stratified design. Statisticians should be consulted in developing the sample design for a biological criteria and monitoring program.

Selecting a Design

The selection of a station array for bioassessment will depend on the nature of the study and/or the desire to delineate the areal extent of impairment. A randomized station selection is most appropriate for environmental status and trends surveys such as conducted by EMAP. However, for specific management decision-making, pre-selected stations placed on a gradient such as distance from of a discharge (sometimes termed "nearfield/farfield") may be more appropriate. This method is a form of model-based design, and more accurately identifies suspected sources of impairment, assesses impacts and monitors recovery.

The number of stations to be incorporated in a study design is most heavily influenced by the available resources. A minimum of three control or reference sites is desired to provide some indication of background variability. The number

of test sites may vary from one to several depending on the purpose of the study. The distance between stations could be decreased; i.e., number of stations increased to partially account for the inefficiency of some sampling gear or, conversely, the distance increased; i.e., number of stations decreased once the data have been evaluated.

Index Period

Most monitoring programs do not have the resources to characterize variability or to assess for all seasons. Sampling can be restricted to an index period when metrics are expected to show the greatest response to pollution stress and when within-season variability is small (Holland 1990). A decision must be made between selecting a sampling period that is representative of the biological community, or one that reflects the worst-case conditions for pollution stress. From the traditional perspective of evaluating pollution impacts in fresh water streams, summer-time low flow conditions are often chosen to assess effects from point source discharges. These flow conditions represent minimal effluent dilution in combination with the natural stressors of low water velocity and high temperature in those constrained environments. In contrast, the effects of nonpoint source pollution on the benthic community are often evaluated following periods of high flow since nonpoint source effects on aquatic communities are largely driven by runoff in the watershed. Estuaries and coastal waters accumulate materials from both nonpoint and point sources in a much more dynamic way and thereby confound the assessment so useful for streams.

In bioassessment strategies involving infrequent sampling, the biologically-optimal period for sampling becomes a major consideration. Periods of instability in community structure, including recruitment of young, natural harsh environmental conditions, changes in food source, and migration of certain target populations are all considerations in conducting these biosurveys. The biologically-optimal period, usually mid-summer and sometimes mid-winter, avoids all of these elements and focuses on the time when communities are most stable. The resource manager or biologist will have to choose between these conditions, or select to cover both, depending on the needs of the study.

5.2.6   Optimizing Sampling

Ferraro et al. (1994, 1989) present a method for quantitatively evaluating the optimum macrobenthic sampling protocol, accounting for sampling unit area, sieve mesh size, and number of replicates (n). Their approach allows managers responsible for designing and implementing estuarine and coastal marine bioassessment programs to answer fundamental questions:

- How large should the sampling unit be?;

- What sieve mesh size should be used?;

- How many replicate samples should be taken?

The procedure calculates the "power-cost efficiency" (PCE), which incorporates both the number of samples (n), the cost (field collection effort and lab effort combined) and the expected statistical power for each alternative sampling scheme. See Chapter 12 for a more detailed discussion of statistical power. The various sampling schemes consist of different combinations of sampling gear, gear area, sieve mesh size, and number of replicates. The method allows determining the optimum among a set of sampling schemes for detecting differences in reference vs. impaired stations when the statistical model is a t-distribution for comparing two means. The optimum scheme can be defined as the least costly one capable of reliably (e.g., $\alpha = 0.5$, $1-\beta = 0.95$) detecting a desired difference in the means of a metric between two stations. The approach can be applied to each metric in a test set of metrics and the results aggregated to determine the optimum protocol.

There are four primary steps in assessing the PCE of a suite of alternative sampling schemes:

1. For each scheme, collect replicate samples at paired reference and impaired stations. The observed difference in metric values between the stations is operationally assumed to be the magnitude of the difference desired to be detected. Alternatively, a percentage of the median (e.g., 20%) for a given metric calculated across reference stations could be set as the magnitude of the difference to be detected. In either case, this difference, divided by the standard deviation, is the "effect size" (ES) of interest.

2. Assess the "cost" (c_i), in time or money, of each sampling scheme i at each station. The cost can include labor hours for sampling,

sorting, taxonomic identification, and recording results.

3. Conduct statistical power analysis to determine the minimum number of replicate samples (n_i) needed to detect the ES with an acceptable probability of Type I (α) and Type II (β) error (e.g., $\alpha = \beta = 0.05$).

4. Calculate the power-cost efficiency (PCE) for each sampling scheme by:

$$PCE_i = (n \times c)_{min} / (n_i \times c_i)$$

where $(n \times c)_{min}$ = minimum value of $(n \times c)$ among the i sampling schemes. The reciprocal of PCE_i is the factor by which the optimal sampling scheme is more efficient than alternative scheme i. When PCE is determined for multiple metrics, the overall optimal sampling protocol may be defined as that which ranks highest in PCE for most metrics in the test set.

Chapter 6
Water Column & Bottom Characteristics

Tiers 1-3 contain active survey and site sampling. Procedures for attaining water column and bottom characteristics are generally the same for each tier. The sampling however, occurs more often over the year. Differences are noted where applicable. Table 6-1 compares the level of effort for each tier. However, agencies will decide which components of each tier will be incorporated into their specific programs, then they will select the level of effort appropriate for their program.

6.1   Salinity, Temperature, Dissolved Oxygen, & pH

Salinity, conductivity, temperature, dissolved oxygen, and pH should be measured at each sampling station using a CTD meter equipped with DO and pH probes. Measurements should be made at 1-m intervals through the water column. In shallow, inshore waters, measurements should be taken at the top, middle, and bottom thirds of the depth. For Tier 3, in some southern waters that undergo significant diel temperature changes, it may be desirable to obtain 24-hour temperature profiles using recording equipment.

6.2   Secchi Depth

Secchi depth is usually measured at the deepest part of the transect or grid. Where the area is classified by depth, Secchi measurements should be made at each station. Readings are obtained with a 40-cm plastic or metal Secchi disk that is either white or is divided into black and white quadrants on a nonstretchable line that is calibrated in

decimeters. The disk is lowered into the water until it disappears from view and the depth is recorded. The disk is then slowly raised to the point where it reappears, with the depth being recorded again. The mean of these two measurement is the Secchi depth. Observations are made from the shady side of the boat, without sunglasses, and as close as possible to the water to reduce glare.

6.3   Depth

Depth should be measured at each station using a calibrated depth sounder. Depth can be read off a meter block when sediment sampling by zeroing the block when the sampler is at the water surface. In shallow, inshore waters, a long stick or weighted line calibrated in decimeters may be used.

6.4   Sediment Grain Size

6.4.1   Estimation of "percent fines" (Tier 1)

Analysis of sediment grain size for Tier 1 assessments can be limited to determining the "percent fines" at each station. A rapid wet sieving technique used in Puget Sound (Eaton 1997) can serve as the basis for this characterization. Materials needed for the procedure include:

▸   standard testing sieve No. 230, 63-μm
▸   50-ml plastic beaker (filled to the brim with sediments is about 79-ml)
▸   100-ml plastic graduated cylinder
▸   water bottle(s) with small outlets

Table 6-1. Water Column & Bottom Characteristics. "Addition" refers to added detail or intensities for a parameter initiated in an earlier tier.

Characteristic	Tier	Collection Method	Indicates
Salinity	1 2 3	-measure at each sampling station, CTD meter -continuous or 1-2-m intervals through water column -shallow/inshore -top, middle, bottom thirds of depth	Distribution of flora and fauna
Temperature	1 2	-measure at each station, CTD meter -1-2-m intervals through water column -shallow/inshore -top, middle, bottom thirds of depth	Rate of chemical reactions and biological processes
Temperature	3 addition	-some southern waters undergo significant diel changes, it may be desirable to obtain 24-hour temperature profiles	
Dissolved Oxygen	1	-measure at each station, CTD meter w/ DO probe -continuous or 1-2-m intervals through water column -shallow/inshore -top, middle, bottom thirds of depth	Possible reason for modified behavior, reduced abundance & productivity, adverse reproductive effects, and mortality
Dissolved Oxygen	2 addition	-measure early in morning at each station at minimum	
Dissolved Oxygen	3 addition	-collect along a depth profile from surface to within 1-m of bottom at 1-2-m intervals -in cases of hypoxic site: recording DO meters may be deployed (EMAP - Louisianian Province, Engle et al 1994)	Resource agency may determine need for more detailed information to diagnose sources & causes of impairment
pH	1 2 3	-CTD w/ pH probe -1-2-m intervals -top, middle, bottom thirds of depth	Chemical condition, pollutant input, high concentrations of phytoplankton
Secchi Depth (Turbidity)	1 2 3	-deepest part of transect/grid -if area classified by depth, measure at each station -See Section 6.2 for complete procedure	Reduction of light penetration, deposition of mud and silt, possible contaminated sediment "hot spots"
Depth	1 2 3	-each station w/ calibrated depth sounder -read off meter block when sediment sampling -shallow/inshore waters: long stick or weighted line calibrated in decimals	Depth at sampling station, possible dredging or sediment loading
Nutrients	2 3	-collected w/ bottle samplers or pump -see Section 6.8 for complete procedure	Nutrient loading
Acid Volatile Sulfides	3	-each station during index period, and any other sampling visits through year -once accurate AVS exists for each station, analytes only performed once per year (during index period)	Bottom characteristics, detailed purposes in Section 3.5.4
Water Column Contaminants	3	Choose One: -USEPA's list of Priority Pollutants, Hazardous Substance, or Target Compound/Analytes -same compounds targeted in EMAP-E (Table 3-1) -develop own list (see Section 6.10 for more detail)	Trace distribution of contaminants from a source or to ID potential sources

Characteristic	Tier	Collection Method	Indicates
Sediment Grain Size	1	-determine "percent fines" at each station, see Section 6.4.1 for complete procedure	Spatial and temporal changes of the benthic habitat, evaluate condition of benthic habitats
	2	-see Section 6.4.2 for complete procedure	Determine extent or recovery from environmental perturbations
	3		Assist in providing early warnings of potential impacts to the estuarine ecosystem
Total Organic Carbon	2	-see Section 6.9 for complete procedure	Provide information regarding sediment organic content (possibly influenced by sewage outfalls)
	3 addition	-measure additional sediment analytes	Examine potential influences of outfalls, ID potential contaminant "hot spots"
RPD Layer Depth	1	-vertical bisection, distance from sediment surface to a noticeable change in color from brownish (oxidizing conditions) to gray (reducing conditions)	Note presence/absence of benthos; learn about life history, taxa abundance, & major taxa biomass distribution; more large, deep dwelling species="healthy" system
Total Volatile Sulfides	1	-deepest section along transect/grid -see Standard Methods (APHA 1992) for sampling & analytical methods	Sediment and carbon content
Sediment Contaminants	1	-conducted at outset of survey	Positive=severe impacts influence spatial sampling design, causal investigations Negative=subsequently collected biological info. essential to ID other, possibly more subtle stresses Provide insight on limiting factors in benthic community
	2	*like TOC, if toxicity tests are initially negative, no need to repeat annually unless biological data from infauna indicate otherwise	
	3	Choose from three approaches: -based on EPA's contaminant lists -NOAA NS&T suite of contaminants (used by EMAP) -targeted list *see Section 6.12 for complete procedure and rationale	

- ▸ small stainless or plastic spatula
- ▸ stainless butter knife
- ▸ hose with nozzle (if running water is available).

Detailed directions for performing this wet-sieving technique are as follows:

Fill a 50-ml plastic beaker to the brim with the sediment to be analyzed. The capacity of the completely filled beaker can be measured using water and the 100-ml graduated cylinder. Clean away any sediment that might adhere to the outside of the beaker. Carefully wash this sediment through a 63-μm standard sieve (USA standard testing sieve No. 230) with stainless steel mesh. The sieve itself is about 9" in diameter with a 2" stainless lip. Be careful not to overflow the sieve with rinsing water. It may be easier to wash half of the sediment through at a time. If running water is available, use a small brass nozzle on the end of the hose with very low water pressure when washing the sediment, otherwise the sediment will need to be washed using the water bottle. If there are occasional large worm tubes or shells, these are discarded and replaced with an approximately equal volume of sediment. The sediment remaining on the sieve is the coarse-grained fraction. This is washed to one side of the sieve, and then carefully placed into the plastic 100-ml graduated cylinder with a stainless steel butter knife, and finally with the small stainless spatula. The water bottle is then used to wash any remaining sediment directly into the graduated cylinder, and to wash down the sides of the cylinder. Let the sediment-water mixture settle in the 100-ml graduated cylinder for approximately 5 minutes until the supernatant water is clear. This may take longer for very fine-grained sediments. Note the volume of the coarse-grained fraction which remains after sieving. This can be divided by the original volume to obtain the percentage of the coarse fraction. The standard usage, however, is for percent fine-grained fraction or "**percent fines**". This is calculated by subtracting the volume of sediment remaining in the cylinder (ml of coarse-grained fraction) from the original volume, and dividing this number (ml of fine-grained) by the original volume to obtain the percent fines.

6.4.2 Sediment Grain Size (Tiers 2 and 3)

Additional grain size data for Tier 2 and Tier 3 assessments should include determination of the size distribution using a standard graded sieve series. This analysis should be performed for a sediment sample collected at each sampling station. In the early years of the assessment program, this analysis should be performed for each sampling period. When an accurate sediment characterization exists for the area of each station, sediment grain size analysis could be performed only annually or biennially (on samples collected in the index period), unless the agency believed that sediment conditions at a site may have changed. This could occur, for example, following a major storm. Buller and McManus (1979) provide a review of the methodological and statistical analysis of sediment samples. If seasonal variations in grain size are exhibited, it is recommended that direct comparisons between samples collected during different seasons be avoided. Studies investigating interannual variation in the percent composition of grain sizes should be conducted during the same season (preferably the same month) each year. Furthermore, it is recommended that grain size be sampled when contaminant concentrations are expected to be at their highest level to evaluate worst-case scenarios.

6.5   RPD Layer Depth

The concept behind using the depth distribution of benthic macroinvertebrates is based on the premise that "healthy" benthic communities in fine sediments in meso- and polyhaline waters consist of relatively large, deep dwelling species; while impaired areas will have fewer of these organisms. The depth distribution of benthic infauna in sediments integrates functional parameters such as life history, taxa abundance, and major taxa biomass distribution.

6.6   Total Volatile Solids

Total volatile solids (TVS) is the Tier 1 indicator for sediment carbon content. TVS should be determined for the deepest station along each transect or grid, based on the assumption that deeper stations will represent sinks for organic carbon in the sediments. Sampling and analytical methods are discussed in Standard Methods (APHA 1992).

6.7   Sediment Contaminant Toxicity

Sediment toxicity testing is a diagnostic indicator for Tier 3. When results are positive for a station, severe impacts at a known locality will influence spatial sampling design and causal investigations. Where toxicity test results are negative throughout the set of stations sampled, subsequently collected biological information is essential to identify other, possibly more subtle stresses on the system.

6.7.1   10-day Static Sediment Toxicity Test with Marine and Estuarine Amphipods

ASTM (1998a) and USEPA (1994b) developed procedures that measure short term adverse effects of potentially contaminated sediment, or of a test material experimentally added to sediment, on marine or estuarine infaunal amphipods during static 10-day exposures for the following species: *Rhepoxynius abronius*, *Eohaustorius estuaris*, *Ampelisca abdita*, *Grandidierella japonica*, and *Leptocheirus plumulosus*. The amphipod *Corophium insidiosum* has also been used in standard testing (Reish and Lemay 1988). Solid phase tests use overlying water in aerated 1-L glass test chambers. Mortality and sublethal effects such as growth, emergence of adults, and inability to bury in clean sediment are determined after exposure of a specific number of amphipods (usually 20) to the test sediment. Response of the amphipods to the test sediment is compared with the response observed in control or reference sediment. The negative control sediment is used to provide a measure of the acceptability of the test by providing evidence of the health and relative quality of the test organisms, the suitability of the overlying water, and test conditions and handling procedures (ASTM 1998b, USEPA 1994b). The reference sediment, which is similar in physical characteristics to the test sediments and typically collected from a similar location, is used as the basis for interpreting data obtained from the test sediments (ASTM 1998b).

The toxicity of field-collected sediments may be assessed by either (a) testing the whole sediment and testing for significant differences in responses between reference or control and test sediment exposed animals or (b) testing dilutions of a test sediment with clean sediment to obtain an LC_{50} or other effect concentration, for survival, reburial success, or growth (ASTM 1998b, Nelson et al. 1993, Swartz et al. 1995).

6.7.2 10-day Static Sediment Toxicity Test with Marine and Estuarine Polychaetous Annelids

Marine or estuarine infaunal polychaetes are used in whole sediment tests during 10-day or 20- to 28-day exposures to determine adverse effects of potentially contaminated sediment, or of a test material added experimentally to sediment. Polychaete species include *Neanthes virens* for the 10-day and *Neanthes arenaceodentata* for the 10-day and 20- to 28-day tests (ASTM 1998c). Other polychaete species that have been used in similar sediment testing include *Capitella capitata*, *Ophrotrocha diadema*, and *Ctenodrilus serratus* (Reish and Lemay 1988). The 10-day test measures effects of contaminated sediment on polychaete survival. The 20- to 28-day test determines effects of contaminated sediment on polychaete survival and growth. If smaller species are used, such as *N. arenaceodentata*, five worms are placed in a 1-L glass test chamber with a minimum sediment depth of 2- to 3-cm and the overlying water is aerated. Either young adults or recently emerged juvenile (2- to 3-weeks post-emergence) worms are used in the 10-day test; only recently emerged (2- to 3-weeks) juveniles are used in the 20- to 28-day test. Survival of worms exposed to the test sediment is compared with the survival in a negative control or reference sediment in either test. If larger species are used, such as *N. virens*, ten worms are placed in a glass aquaria (4- to 37-L) with a minimum sediment depth of 10-cm and the overlying water is aerated.

The percent survival of polychaetes exposed to field-collected sediment is compared to those exposed to a negative control or reference sediment in 10-day tests. Survival and body weight of surviving animals is compared to those exposed to negative control or reference sediment in 20- to 28-day tests. The toxicity of field sediments may also be assessed by testing dilutions of highly toxic test sediments with clean sediments to obtain either an LC_{50} or other effect concentration of the material.

6.7.3 Static Acute Toxicity Tests with Echinoid Embryos

Echinoderm embryos and larval form sea urchins (*Strongylocentrotus purpuratus* and *Strongylocentrotus droebachiensis*) and sand dollars (*Arbacia punctulata*, *Lytechinus pictus*, and *Dendraster excentricus*) have been used in marine sediment interstitial (pore) water tests (ASTM 1998a). Interstitial water from marine sediments is isolated using either *in-situ* peepers (Sarda and Burton 1995, Brumbaugh et al. 1994, Bufflap and Allen 1995), suction in the field (Watson and Frickers 1990), laboratory centrifugation (Ankley et al. 1991, Burgess et al. 1993, Kemble et al. 1994, ASTM 1998b), or sediment squeezing (Long et al. 1990). Embryos are obtained by inducing adults to spawn, using either physical (e.g., electric stimuli) or chemical (injection of potassium chloride) means, and then combining gametes.

Embryos are exposed to the test pore water and controls (culture water) for 48- to 96-hours, depending on the species and test temperature. The test measures the proportion of embryos or larvae that develop into normal pluteus larvae. Pore waters can be tested "whole"; i.e., undiluted, and organism responses expressed in terms of a significant difference between controls and test waters. Alternatively, pore water samples can be diluted with known, clean culture water and the results expressed as an LC_{50} or other

effect concentration with confidence limits.

6.7.4 Toxicity Tests Using Marine Bivalves

Juveniles of the marine bivalve species, *Mulinia lateralis*, have been used in whole sediment tests (Burgess and Morrison 1994). Juveniles are exposed for 7-days to determine adverse effects of potentially contaminated sediment, or of a test material added to sediment. Bivalve responses measured include survival and growth, (total organism dry weight). Ten juvenile bivalves (four weeks old) are placed into six replicate chambers per sediment or treatment. The sediment exposure chambers are prepared by placing approximately 1.0-cm deep sediment into 150-ml dishes, followed by the addition of 100-ml filtered 30-gkg^{-1} seawater. Upon initiation of the test a subsample of organisms are set aside for determination of initial juvenile weights. Bivalve survival in test chambers is compared to survival of bivalves in the negative control or reference sediment. Dry weight of the surviving organisms in test chambers is compared with dry weight of surviving organisms in the reference sediment, and to the dry weight of the subsample set aside at the initiation of the test to determine growth.

Similar to echinoderm testing summarized in Section 6.7.3, bivalve larvae have also been used in sediment pore water and sediment elutriate toxicity tests. Species used include *Crassostrea gigas* and *Mytilus edulis* (PSEP 1995). Bivalve larvae are obtained from laboratory-cultured adult brood stock, which are induced to spawn. Developing embryos are exposed to the pore water or elutriate at 20°C for 48-60-hours using static-test conditions. At test termination, subsamples of the larvae are examined and the percentages of mortality and abnormal survivors are determined and analyzed.

6.8 Nutrients (Tiers 2&3)

Water column samples for nutrient analysis can be collected using bottle samplers such as Kemmerer, Van Dorn, Niskin, or Nansen samplers. A pump may be used as an alternative sampling device. In shallow water less than 2-m depth, a mid-depth sample at each station should be obtained for nutrient analysis. In waters greater than 2-m depth, samples should be collected at each station at 1-m below the surface, 1-m above the bottom, 1-m above the pycnocline, 1-m below the pycnocline, or at mid-depth. Analytical methods for NH_4-N, NO_3-N, NO_2-N, Kjeldahl nitrogen, total N, and total and reactive P; i.e., ortho-P, are presented in APHA (1992) and USEPA (1994c). These nutrient analyses will help identify eutrophication factors affecting biocriteria development, as well as supplement the USEPA's nutrient criteria initiatives so that multiple objectives can be accomplished at once.

6.9 Total Organic Carbon (Tiers 2&3)

In Tier 2, the primary purpose of measuring total organic carbon (TOC) is to provide information regarding sediment organic content, which might be influenced by sewage outfalls containing high organic levels. As noted in Chapter 3, TOC in the sediment is an important analyte for the purpose of evaluating the bioavailability of organic pollutants and metals adsorbed by sediments or contained in sediment porewater. Data on sediment TOC collected in this tier can be used to examine potential influences of outfalls in addition to potential sediment contaminant "hot spots" that can be

assessed in Tier 3 with the measurement of additional sediment analytes.

Standard methods for TOC analysis are presented in APHA (1992). In the early years of the assessment program, TOC analysis should be performed for each station in each sampling period. Once the resource agency is confident that an accurate characterization of sediment TOC exists for each station, the analysis could be performed only once every two or more years (on samples collected in the index period), unless stations that appear to be influenced by organic input (e.g., sewage outfalls) are identified. In this case, TOC analysis should continue to be performed for each sampling period for these stations.

6.10 Water Column Contaminants (Tier 3)

Water column contaminants such as organic compounds (e.g., herbicides, pesticides, hydrocarbons) and metals may be important indicators of sources and causes of impairment to biological assemblages in estuaries and coastal marine waters. Decisions on which chemicals to include in Tier 3 assessments can be difficult. Three approaches to selecting contaminants might be useful. One approach would be to analyze for all chemicals listed on USEPA's Priority Pollutant, Hazardous Substance, or Target Compound/ Analyte Lists. A second approach would be to analyze for the same compounds targeted in the EMAP-Estuaries program (refer to Table 3-1). A third approach would be to develop a targeted list. In this latter approach, the historical information from Tier 0 and subsequent follow-up inquiries of land use in the suspect area could point to common pesticides, herbicides, or industrial products or byproducts that could form the basis of a select list of contaminants to analyze. Sources for

this information also include NPDES permit records and discharger toxicity test results. In any case, three replicate water samples should be collected at each sampling station within an appropriate index period and on at least three other visits during the year to capture temporal variations in contaminant concentrations. Historic water contaminant data, plus data collected in this tier, can be used by the state to determine a more limited list of analytes for subsequent years of the assessment and biocriteria program.

The same type of sampling bottle used to collect water samples for nutrient analysis may be used for contaminant samples. USEPA (1992) and APHA (1992) contain detailed information on analytical methods.

6.11 Acid Volatile Sulfides (Tier 3)

Details of the purposes for measuring acid volatile sulfides (AVS) present in bottom sediments are provided in Section 3.5.4. Given the diagnostic intent of a Tier 3 assessment, it is important to include this analyte in determinations of bottom characteristics only if metals are suspected as a cause of biological degradation. Allen et al. (1993) discuss analytical methods for this parameter. AVS measurements should be made on sediment samples collected at each station during an appropriate index period and any other sampling visits made throughout the year. Once the resource agency is confident that an accurate characterization of sediment AVS exists for each station, the analytes should be performed only once per year (on samples collected in the index period).

6.12 Sediment Contaminants

As with water column contaminants, three approaches to selecting analytes could be used: (1) a full scan based on USEPA's contaminant lists; (2) the NOAA National Status and Trends suite of contaminants used by the EMAP program (refer to Table 3-1); or (3) a targeted list.

In this latter approach, the historical information from Tier 0 and subsequent follow-up inquiries of land use in the suspect area could point to common pesticides, herbicides, or industrial products and byproducts that could form the basis of a select list of contaminants to analyze. In addition to sampling organisms for contaminants, sediment samples should be collected from the device used for sampling benthic infauna. The surface sediment (top 2-cm) should be removed from replicate grab samples and composited. During collection, care should be taken to avoid collecting material from the edge of grabs and to use only samples that have undisturbed sediment surfaces. The composite sample should be homogenized, and a subsample taken for measurement of contaminant concentrations. Analytical methods are discussed in APHA (1992) and USEPA (1992).

Chapter 7

Tier 0: Desktop Screening

The breakdown of screening and sampling in the following chapters that focus on the Tiered Approach are just one way of designing a state-wide monitoring program. Agency analysis of resources and program objectives should direct the custom development of any monitoring program.

The desktop screening assessment (or Tier 0) consists of compiling documented information for the estuary or coastal marine areas of concern through a literature search and sending survey questionnaires to local experts. No field observations are made at this assessment level. Desktop screening should precede any of the three subsequent tiers. Its fundamental purpose is to support the planning for monitoring and more detailed assessments. It incorporates time and cost efficiencies, allowing evaluation of a

large number of sites, and identifying potentially affected areas for further investigation in higher tiers. Table 7-1 gives an overview of the components, sources, and uses of a desktop screening assessment.

7.1 Area and Geomorphometric Classification

The size and classification of the estuary indicates the potential for the environment to respond to various types of impacts. In addition, the classification refers to the type of circulation (e.g., gravitational, tidal, wind-induced) that dominates the estuary. Well-recognized estuary types include:

- ▸ Coastal plain estuary;
- ▸ Lagoon;

Table 7-1. Tier 0 Desktop screening for estuaries and coastal marine waters.

Component	Information Source	Use
Estuary area	USGS quad maps, GIS	-support planning for monitoring and more detailed assessments -incorporates time & cost efficiencies -allows evaluation of a large number of sites
Geomorphic classification	USGS quad maps, GIS	
Habitat type	NOAA bathymetry charts; historic surveys by federal, state agencies, and universities	
Biological assemblages	Historic data from federal, state agencies, and universities. NMFS for marine mammal data	
Watershed land use	USGS land use maps; state and county planning agencies; local zoning agencies; USDA CSREES	
Population density	US census data	
NPDES discharges	State water quality agency and regional USEPA offices, PCS database	
Water column & bottom characteristics	Historic data from federal, state agencies, and universities; STORET, NODC databases	

- ▸ Fjord;
- ▸ Tectonically-caused estuary.

7.2   Habitat Type

Partitioning of the resource by habitat type (open water, soft bottom substrates, hard bottom substrates, aquatic macrophytes, high/low energy beaches, sandflat, mudflat, emergent marsh) will usually be required and the extent of the partitioning will depend on the size of the system and environmental gradients. Initial subdivisions should be based on salinity gradients, water depth, and sediment type, particularly in coastal marine areas.

7.3   Watershed Land Use

The pollutant and sediment load of fresh water inflow into the estuary will inevitably have some form of impact on habitat and biota and this land use information may subsequently help identify causes of impairment. Nonpoint source pollution has been shown to be a major contributor to the degradation of our aquatic resources. Land use information will help determine the type of contaminants that are being flushed into the estuary. For example, storm water runoff from urbanized and industrial areas may contain various types of toxins. Runoff from agricultural areas could be expected to contain fertilizers, pesticides, and sediment. Fertilizers have the potential to accelerate eutrophication by excessive nutrient enrichment, while pesticides may have at least short-term toxic effects.

7.4   Population Density

This indicates the potential for the whole array of impacts to the estuary and coastal marine waters from concentrated human activity. The more populated the area surrounding the

estuary or coastal region, the higher the potential for human-induced impacts.

7.5   NPDES Discharges

Industrial and municipal point source dischargers must file monthly discharge monitoring reports (DMRs) that provide the effluent concentrations for the contaminants in the effluent which they are required to monitor. This data is accessible via USEPA's PCS. Knowing the number, type and location of point source dischargers could provide the background information necessary for characterizing the contaminants entering the estuary and the regions within the estuary or coastline that would be most affected by the discharge.

7.6   Biological Assemblages

Existing information on any of the target biological assemblages (benthos, fish, macrophytes, photoplankton, zooplankton, epibenthos, paleoenvironmental systems) can be valuable for:

- ▸ Identifying potential reference sites, and potentially impaired areas;

- ▸ Determining presence/absence of major taxonomic groups and indicator organisms;

- ▸ Evaluating spatial and temporal variability of the biological assemblages.

This information can be used to help determine target assemblages for higher-level tiers and the sampling design and methods that might be appropriate.

7.7 Water Column and Bottom Characteristics

Existing data on water column and bottom characteristics will be crucial to support the identification of appropriate sampling strata based on salinity, grain size, or depth. Further, this information can help states identify potentially impaired areas; i.e., areas receiving high nutrient loadings or containing contaminated sediments.

Chapter 8
Tier 1

The Tier 1 assessment is just one way of completing a minimal biological assessment or simple field screening. Specific agency needs will ultimately decide the components of any state monitoring program. The time period of sampling should be selected to allow states to answer the question: "What information do we want to obtain from a single site visit?" For example, it could be conducted from a single field visit during late summer when low dissolved oxygen concentration, due to stratification and eutrophication, is most likely to occur or during some other chosen index period, depending on the monitoring purpose. It builds on the information compiled in the desktop screening assessment and consists of sampling one or more biological assemblages and collecting data on water column and bottom characteristics (Chapters 5 and 6). Tier 1 might roughly identify whether an estuary or coastal marine waters are nutrient enriched and can distinguish among broad probable causes if the nutrient state is different from expectations (reference conditions). This assessment tier enables:

- ▸ coarse identification of nutrient state based on chlorophyll *a* concentration, and identification of point and nonpoint probable cause if stations are carefully selected and spaced;

- ▸ detection of emergent wetlands and shore zone fish habitat loss from shore zone survey and macrophyte assessment;

- ▸ detection of loss of submerged aquatic macrophytes;

- ▸ detection of potential impairment of benthic macroinvertebrate and fish assemblages;

- ▸ detection of oxygen stress.

The Tier 1 assessment will not allow separation of multiple probable causes. It can establish the initial habitat classification scheme and identify several possible causes of impairment, including point sources, nearfield nonpoint sources (in the immediate shore zone of the coast or estuary), and farfield nonpoint sources (from land use in the drainage). It cannot, however, identify the most probable from among several possible causes. It should also help establish the most likely sites to use in developing the reference condition and test their candidacy for this preliminary phase of biocriteria development. Table 8-1 gives an overview of the components, data collection methods, and indicators for Tier 1.

8.1 Benthos

Sampling and analysis of benthic infaunal macroinvertebrates in Tier 1 is intended to provide a rapidly obtained snapshot of the condition of the benthic assemblage. It is recognized that this assemblage, and the methods presented, will be most appropriate for sites with soft sediments (e.g., mud, silt, sand). For sites with hard bottom substrates, other biological assemblages (e.g., fish, macrophytes, phytoplankton) could be selected to provide information on the biological condition of the target waters.

The sampling strategy presented here consists of collecting replicate grab

Table 8-1. Tier 1 Assessment. Requires single field visit in spring or summer index period.

Component	Data Collection	Indicator of	Uses
Biological Assemblages			
Benthic Infauna	*3 replicate grabs *x-section of Smith-McIntyre or Young grab *measure RPD depth *brief description of classes & families of benthos present in grab *record faunal presence/absence of benthos above and below the RPD depth in sediment x- section	DO stress, toxicity, low productivity, nutrient enrichment, habitat impairment	-Identification of nutrient state based on chlorophyll *a* concentration & identification of point and nonpoint probable causes (if stations are carefully selected & spaced) -Detection of emergent wetland & shorezone fish habitat loss from shorezone survey and macrophyte assessment -Detection of loss of submerged aquatic macrophytes
Fish	*3 trawls *3 seines *species counts *measure standard lengths *record external abnormalities	DO stress, toxicity, habitat impairment	
Macrophytes	*aerial photos (if possible - or estimate from shorezone survey *% cover estimate *record dominant taxa	Nutrient enrichment, sediment loading	
Phytoplankton	*chlorophyll *a* *record blooms *identify dominant species	Nutrient enrichment	
Water Column Characteristics			
	*salinity/conductivity *temperature *DO *pH *Secchi depth *depth *TSS for seagrass	DO stresses, eutrophication, stratification, turbidity	-Detection of potential impairment of benthic macroinvertebrate & fish assemblages
Bottom Characteristics			
	*grain size estimate & description *RPD layer depth *TVS *sediment toxicity	Habitat modification, DO stress, sediment toxicants	-Detection of oxygen stress

samples at each sampling site, taking a vertical cross-section of the sample, and measuring the RPD layer depth to record the presence/absence of benthos above and below the RPD depth in the sediment cross-section. In addition to the actual presence of organisms, evidence of their presence, such as bivalve siphons, siphon impressions in clay/mud, or polychaete burrows, should also be noted. The investigator may also wish to sort and identify the organisms found above and below the RPD depth for additional information relative to Tier 2. The method presented here is a simplification of the Benthic Assessment Method developed by Diaz and Nelson (1993). Functional attributes of the benthic infaunal community that can be evaluated using these procedures include:

▸ *Species Life Histories* The presence of relatively large and long-lived species, especially those found deeper in the sediments, indicate higher quality habitat than does the presence of small and short-lived taxa;

▸ *Major Taxa Abundance* High abundance of only a few taxa, usually pollution tolerant ones, indicates a degraded environment;

▸ *Major Taxa Biomass Distribution* Larger organisms, hence a higher biomass per individual, are more prevalent in better quality habitats;

▸ *Vertical Distribution of Biomass* Organisms living below 5-cm in soft substrates indicate a relatively high quality habitat.

8.1.1 Sampling Procedure

The primary objective of benthic infaunal macroinvertebrate sampling in Tier 1 is to determine whether there are any large organisms below the RPD depth. The recommended sediment sampling procedure involves collecting three replicate grab samples at each station using a Smith-McIntyre or Young grab. The selection of sampling gear should be made to maximize compatibility with historic data. For example, the state of Texas uses an Ekman grab, and has an approximately 25-year data record using this gear type. The sediment sample is vertically bisected using a sheet metal partition. The RPD layer depth is noted and measured, if present, as the distance from the sediment surface to a noticeable change in color from brownish (oxidizing conditions) to gray (reducing conditions). The sediment above the RPD depth is removed and wet-sieved separately; the remaining portion of the sample is also wet-sieved.

A sieve with mesh size appropriate for the region should be used. The presence or absence of benthic infauna in either subsample is noted. If present, the classes and families should be noted and recorded.

8.1.2 Index Period

Benthic infaunal macroinvertebrates are sampled once during an appropriate index period, the timing of which is driven by the goals of the Tier 1 assessment and regional considerations.

8.1.3 Analysis

Note the presence/absence of an RPD layer and any infauna (or evidence of infauna) below 5-cm depth in the sample. If present, identify benthic infauna to class and family and record abundance.

8.2 Fish

A Tier 1 assessment of the fish assemblage is intended to provide a rapid evaluation of its presence and overall composition. Fish sampling in Tier 1 can include shallow-water, pelagic, and demersal fish communities (Carmichael et al. 1992, Eaton and Dinnel 1994, Guillen 1995a).

8.2.1 Sampling Procedure

Various nets can be used to sample littoral and sublittoral areas. It is recommended that trap nets (gill or fyke nets) be set and fished twice a day for 2- to 5-days. Due to the risk of boating mishaps and vandalism, it is recommended that investigators stay with the nets while they are being fished. Fish sampling methods are detailed in Klemm et al. (1992).

▸ Gillnets are set in littoral areas at right angles to the shore or to

longshore fish movement. Gillnets usually extend into sublittoral areas. Smaller mesh size (0.5") is used in shallow areas and up to 2 to 2.5" mesh is used further away from shore. To reduce size selectivity, an experimental gillnet consisting of panels of five different mesh sizes is commonly used;

▸ Trawl nets and sonar can be used to sample pelagic and demersal areas. The length of the towline (warp) should be at least six times the depth of water and a trawl speed of about 2-knots over a 0.5-nautical mile distance is appropriate for coastal marine waters. These values of warp length and trawling distance can be reduced in estuaries. A 20-ft trawl (16-ft effective trawl mouth) is appropriate in marine waters, but an 8- or 10-ft trawl is easier to tow in restricted waters.

8.2.2 Sample Processing

Sampling duration and area or distance sampled (from DGPS) are recorded in order to determine sampling effort. Species are identified and enumerated. Fish should be carefully removed from the net to avoid undue handling and damage. The catch should be sorted by species, and length measurements made of each individual. This measurement is usually total length, but fork length or standard length can also be used. At the time of measurement, any deformities, ulcerations, bleeding, fin rot, bulging eyes or other disease indicators should be noted and those fish saved for histopathology. It is important to distinguish net damage from pre-existing conditions, if possible. Wet weights can be taken by species by weighing the fish either individually using a platform scale or collectively from tared hanging scales, depending on the number of fish caught. As a

small matter of convenience, both scales should weigh in metric units. Those animals not saved for further

examination should be promptly returned to the water.

The investigator should consult with State and University fish pathologists of the region for those most appropriate sample preparation and preservation techniques. Usually iced or frozen specimens are inappropriate and in some cases formaldehyde or other tissue preservatives must be carefully used if meaningful samples are to be presented. Generally, small fish can be tagged and placed whole in 10% formalin. Larger fish will require dissection in the field and the tissue samples tagged and preserved in the same manner. Protocols for preservation and dissection should be obtained from the laboratory/fish pathologist that will receive the samples.

When collected, reference specimens of each species from each site are preserved in 10% formalin in a labeled jar and retained by the state ichthyological museum or other designated repository to constitute a biological record. This is especially important for uncommon species, for species requiring laboratory identification, and for documenting new distribution records. Later, all specimens should be transferred from formalin to 70% alcohol for long-term storage.

8.3 Macrophytes

Areal coverage and distribution of submerged aquatic macrophytes is estimated from aerial photographs, if available, and ground-truthed at the site. The dominant taxa may be field-identified from vegetation samples collected in shallow waters. Detailed

macrophyte monitoring and assessment procedures are included in USEPA (1992), Ferguson and Wood (1994), and Orth et al. (1993).

8.4   Phytoplankton

Phytoplankton standing stock is estimated by chlorophyll *a* measurements. A sample is collected at each station at one-half the Secchi depth using a Kemmerer or Van Dorn sampler. Chlorophyll *a* is determined using a fluorometer or spectrophotometer as discussed in APHA (1992). The presence of any phytoplankton blooms observed during the cruise should be noted. Dominant phytoplankton species should be identified.

Chapter 9
Tier 2

Tier 2 assessment is considered a routine biological survey that incorporates two or more field visits per year to capture variations due to seasonal differences. Tier 2 comprises increased sampling effort and additional assemblages compared to Tier 1. It includes two or more biological assemblages (benthos, fish, macrophytes, phytoplankton, or epibenthos), in 2 or more visits per year, in addition to more detailed characterization of the water column and bottom (Chapter 5). State agencies can modify this schedule to accommodate their program objectives.

This level is sufficient for identification of appropriate habitat classes and determination of the reference condition for development of biological criteria. Data collected in Tier 2, which incorporates both Tier 1 and Tier 0, should permit the state to confidently develop biocriteria and apply them to identify problem areas. This assessment level enables:

▸ Establishment of the biocriteria "benchmarks" for decision-making about impaired areas; including identification of priorities;

▸ Identification of trophic state based on chlorophyll *a* and water column nutrient measurements;

▸ Detection of impairment of benthic macroinvertebrate, fish, or epibenthos assemblages, and evaluation of potential causes of the impairment;

▸ Measurement of extent of macrophyte coverage;

▸ Identification of phytoplankton taxa responsible for blooms.

A Tier 2 assessment should allow identification of multiple probable causes of impairment, given an adequate number and placement of sampling stations. This includes point sources, and nearfield and farfield nonpoint sources. Preliminary management plans in response to the biocriteria information can be developed. Table 9-1 gives an overview of the components, data collection, methods, and indicators for Tier 2.

9.1 Benthos

Sampling and analysis of benthic infauna in Tier 2 is intended to provide a level of assessment consistent with routine benthic macroinvertebrate surveys presently conducted by states in estuaries and coastal marine waters. As with Tier 1, this assemblage, and the methods presented, will be most appropriate for soft sediments. For sites with hard bottom substrates, other biological assemblages (e.g., fish, macrophytes, phytoplankton) could be selected to provide information on the biological condition of the target waters.

The sampling strategy for Tier 2 entails a minimum of two field collection visits, one of which should occur within the chosen index period. Organisms are identified to genus and species to determine major taxa and the presence of indicator organisms. Water column and bottom characteristics are also measured to evaluate the status of physicochemical conditions.

Table 9-1. Tier 2 Assessment. Requires two or more field visits, one of which should occur within chosen index period. In addition to requirements from Tiers 0 & 1.

Component	Data Collection	Indicator of	Uses
Biological Assemblages			
Benthic Infauna	*3 replicate grabs (or as determined by testing; see Section 5.2.6 *identify major taxa and indicator spp. of each grab to genus and species		-Establishment of biocriteria "benchmarks" for decision-making about impaired areas, including identification of priorities -Identification of *trophic* state based on chlorophyll *a* and water column nutrient measurements -Detection of impairment of benthic macroinvertebrate, fish, or epibenthos assemblages and evaluation of potential causes of impairment -Measurement of extent of macrophyte coverage -Identification of phytoplankton taxa responsible for blooms
Fish	*3 or more replicates *biomass by species		
Macrophytes	*area *maximum depth *identify taxa and measure wet weight of 2-3 samples per transect	Habitat impairment	
Phytoplankton	*identify dominant species, including "nuisance" taxa, on a seasonal basis		
Epibenthos (developmental)	*mid-summer or growing season average at genus and species level *calculate sensitivity metric		
Water Column Characteristics			
	*nutrients: NH_4, NO_3, NO_2, Kjeldahl N, total and reactive P		
Bottom Characteristics			
	*grain size measurements *TOC	Organic enrichment	

9.1.1 Sampling Procedure

Primary objectives of Tier 2 benthic infaunal sampling are to evaluate potential impairment to this assemblage and to generate the data necessary to develop biocriteria. This Tier, unlike Tier 1, incorporates multiple sampling visits to allow a basic discrimination of seasonal differences in the benthic infaunal macroinvertebrate assemblage. The recommended sediment sampling

procedure involves collecting three replicate grab samples at each station using a Smith-McIntyre or Young grab. Sampling gear should be selected to maximize compatibility with historic data. The sediment sample is vertically bisected using a sheet metal partition. The RPD layer depth is noted and measured, if present, as the distance from the sediment surface to a noticeable change in color from brownish (oxidizing conditions) to gray

or black (reducing conditions). The sample should be wet sieved through a sieve mesh size determined to be appropriate for the region (Section 6.3.2). For cost and effort savings, an appropriate diameter subcore (2.5- or 5-cm) can be taken from each of the four quadrants of the intact core. These subcores should be compared to organism counts taken from full cores to establish the baseline relationship between the two. Organisms and sediment fractions should be placed in tagged and labeled sample jars with a 10% solution of magnesium chloride or magnesium sulfate to narcotize the animals. After at least 30-minutes, concentrated formaldehyde with rose bengal dye can be added to the jars to make a 10% solution of formaldehyde by volume. The sediment/organism material should never exceed half the container volume to ensure adequate mixing and fixation of the sample. For preservation, the samples should be transferred to 70% ethanol (APHA 1992).

9.1.2 Index Period

Benthic infaunal macroinvertebrates are sampled once during an appropriate index period, the timing of which is driven by the goals of the Tier 2 assessment and regional considerations. At least one other sampling visit is made outside the index period to capture basic seasonal differences in the assemblages. The timing of this visit(s) will depend on the specific goals of the assessment.

9.1.3 Analysis

Organisms in each sample are identified to genus and species. Metrics selected by the state can then be calculated to assess the condition of the assemblage. Metric values can then be used to help develop biocriteria against which the condition of the macroinvertebrate assemblage is evaluated.

9.2 Fish

Tier 2 assessment of the fish assemblage is intended to provide data sufficient to evaluate impairment and to develop biocriteria. Fish sampling in Tier 2 can include shallow water, pelagic, and demersal fish communities (Carmichael et al. 1992, Eaton and Dinnell 1994, Guillen 1994).

9.2.1 Sampling Procedure

See Section 8.2.1 for full procedure on sampling fish.

9.2.2 Sample Processing

See Section 8.2.2 for full procedure on sample processing.

9.2.3 Analysis

Based on the enumerated species list, metrics selected by the state can be calculated to evaluate potential impairment to the fish assemblage and to develop biocriteria for this assemblage.

9.3 Macrophytes

Tier 2 assessment of macrophytes is intended to provide sufficient data to assess impairment to the macrophyte assemblage as a significant habitat variable and potential element of biocriteria. Because of its importance as habitat for other assemblages, procedures for Tier 2 assessment of macrophytes are considerably more involved than for Tier 1.

9.3.1 Sampling Procedure

The extent of coverage and distribution of macrophytes should be determined

from aerial photographs. Existing aerial photographs are inexpensive; however, they may not be sufficiently recent to depict present macrophyte distribution in the water body. If new aerial photographs are determined to be needed, states should recognize that overflights can be expensive and complicated; often requiring assistance from firms specializing in aerial photography. Factors to consider when planning new overflights include: tidal stage; weather conditions; time of day; and water turbidity (USEPA 1992). Ferguson and Wood (1994) and Orth et al. (1993) describe details of planning aerial overflights, obtaining imagery, photointerpretation, and preparation of macrophyte distribution maps.

A key aspect of interpreting aerial photographs is the performance of ground surveys that serve to confirm the existence of macrophyte beds identified in the photographs, as well as beds that may not be visible in the photos (Orth et al. 1993). Transects can be plotted across macrophyte beds in the various salinity zones within an estuary or within the sampling strata used for marine waters. At each station on the transect a 1-m^2 quadrat can be used for the purpose of measuring percent cover and collecting macrophyte samples for taxonomic identification and measurement of wet weight (USEPA 1992). Depth at the channel-ward or seaward edge of macrophyte extent should be recorded.

9.3.2 Index Period

Aquatic macrophytes should be sampled once during an appropriate index period, preferably during the time of year when they would be expected to be most dense and extensive. Other sampling periods should be selected based on the specific goals of the Tier 2 assessment, perhaps to measure seasonal periods of stress or

diminishment of important nursery or food areas.

9.3.3 Analysis

Percent cover and area may be derived from analysis of aerial photographs. The maximum depth of occurrence is a good indicator of water quality. Taxonomic identification from the field trips will allow development of a species list.

9.4 Phytoplankton

9.4.1 Sampling Procedure

Phytoplankton standing stock is estimated by chlorophyll a measurements. One approach might be three replicate samples collected at each station at one-half the Secchi depth using a Kemmerer or Van Dorn sampler. Another approach would collect a depth-integrated sample through the entire photic portion of the water column. Chlorophyll a is determined using a fluorometer or spectrophotometer as discussed in APHA (1992). The presence of any phytoplankton blooms should be noted. In addition to chlorophyll a measurements, samples from each station should be preserved for subsequent analysis to identify the dominant taxa and those taxa that might be responsible for observed blooms (USEPA 1992).

9.4.2 Index Period

Phytoplankton populations can vary rapidly over space in response to tides and currents, and over time in response to ambient temperature and nutrient inputs. For Tier 2, phytoplankton should be sampled at least once during an index period (usually summer) and at least once outside that index period.

9.4.3 Analysis

Chlorophyll *a* measurements can be used to estimate phytoplankton standing stock. Assuming that chlorophyll *a* is about 1.5% of the ash-free dry weight of algae, algae biomass can be estimated by multiplying the chlorophyll *a* content by a factor of 67 (APHA 1992). This information can be used in concert with the identification of dominant taxa and "nuisance" taxa to assess the overall condition of the phytoplankton assemblage.

9.5   Epibenthos (Developmental)

Although its use as an indicator of estuarine and coastal marine biological condition is considered to be under development, epibenthos could be selected as one of the biological assemblages for a Tier 2 assessment and has potential as an element of biological criteria consistent with fish and benthic invertebrates.

9.5.1   Sampling Procedure

Farrell (1993a, b) describes the use of a beam trawl to collect epibenthos. A beam trawl is a conical-shaped net, open at the large end, which is towed over the substrate surface. The net is kept open by attaching each end of it to a rigid pole or beam. This beam replaces the doors of an otter trawl and forward movement of the boat is not required to keep the net open. The net is constructed in two parts. The body is nylon bolting cloth (50 openings/cm^2), tapering to a plankton net fitted with a removable container. An effective swath width of 1.25-m has been tested in Florida waters (Farrell 1993a, b). In wadeable water, a D-frame net could be used to collect epibenthos, or the beam trawl could be pulled by hand. A relatively short tow length of the beam trawl (4-m,

effectively sampling 5-m^2 of bottom) in estuaries may be beneficial for reducing the sample size and detrital bulk. If a D-frame net is used, at least an equivalent area should be sampled. In offshore waters, it may be necessary to increase the tow length due to reduced organism densities. Small otter trawls or an epibenthic sled sampler can also be used.

9.5.2   Index Period

Epibenthos should be sampled once, preferably during an appropriate index period. For many temperate areas of the country, this is probably mid-summer. Other sampling periods should be selected based on the specific goal of the Tier 2 assessment.

9.5.3   Analysis

The samples should be identified to genus and species. The Farrell Index (described in Chapter 13 - Case Studies, as modified to reflect tolerance values of taxa in the area sampled) should be calculated to provide an assessment of the condition of the assemblage in response to organic pollutants and eutrophication. Other metrics could be calculated based on the specific taxa present.

Chapter 10

Tier 3

Tier 3 is the most rigorous of the assessment tiers. It includes more detailed assessment procedures that allow monitoring agencies to focus on specific water and sediment quality problems in estuarine or coastal marine waters. Tier 3 is intended to provide definitive information needed to act on biocriteria and to measure potential success or failure of the management effort. It allows states to conduct a detailed diagnosis of the sources and causes of impairment to biological assemblages and the physicochemical environment and to monitor their response to subsequent mitigation actions. However, the Tier 3 approach can be customized to accommodate specific state program objectives. Table 10-1 gives an overview of the components, data collection methods, and indicators for Tier 3.

Tier 3 assessments include multiple sampling visits per year (four or more) that occur within each season including the index period. Data collected in Tier 3, which includes information compiled in Tier 0 desktop screening and comprises the information collected in Tiers 1 and 2, involves sampling and measurement of three or more biological assemblages (benthos, fish macrophytes, phytoplankton, zooplankton, or epibenthos), in addition to more detailed characterization of the water column and bottom. A Tier 3 assessment enables:

▸ Identification of nutrient state based on chlorophyll *a* and water column nutrient measurements;

▸ Detection of impairment of benthos, fish, macrophytes, phytoplankton, zooplankton, epibenthos, or paleoenvironmental systems;

▸ Diagnosis of specific sources and causes of impairment;

▸ Measurement of extent of macrophyte coverage;

▸ Identification of phytoplankton taxa responsible for blooms;

▸ Evaluation of seasonal dynamics of biological assemblages;

▸ Detailed monitoring of sites requiring management initiatives to meet the biocriteria;

▸ Inferences of past conditions as a site-specific reference.

10.1 Benthos

Sampling and analysis of benthic infaunal macroinvertebrates in Tier 3 is intended to provide a diagnostic level of assessment. This assemblage, and the methods presented, will be most appropriate for soft sediment. For sites with hard bottom substrates, other biological assemblages (e.g., fish, macrophytes, phytoplankton, zooplankton) could be selected to provide information on the biological condition of the target waters.

The sampling strategy for Tier 3 entails a minimum of four field collection visits per year, one of which should occur within the chosen index period. The remaining visits should occur throughout the year to allow evaluation

Table 10-1. Tier 3 Assessment. Requires four or more field visits, one of which should occur within the chosen index period. In addition to requirements from Tiers 0-2.

Component	Data Collection	Indicator of	Uses
Biological Assemblages			
Benthic Infauna	*determine biomass *calculate multiple metrics		-Identification of *nutrient* state based on chlorophyll *a* & water column nutrient measurements -Detection of impairment of benthos, fish, macrophytes, phytoplankton, zooplankton, epibenthos, or paleoenvironmental systems -Diagnosis of specific sources & causes of impairment -Measurement of extent of macrophyte coverage -Identification of phytoplankton taxa responsible for blooms -Evaluation of seasonal dynamics of biological assemblages -Detailed monitoring of sites requiring management initiatives to meet the biocriteria -Integrate conditions over broad spatial scales
Fish	*5 or more replicates *histopathology on representative subsample of catch	Fishing pressure, disease	
Macrophytes	*stem counts *biomass *record pathology	Toxicity, habitat impairment, disease	
Phytoplankton	*full community characterization to species		
Paleoenvironmental Systems (developmental)	*2 or 3 cores from basin (one-time sample)	Past conditions	
Epibenthos (developmental)	See Tier 2		
Zooplankton (developmental)	*identify to species	Water quality impairment, DO stress	
Water Column Characteristics			
	*pesticides, herbicides *metals	pesticides/ herbicides, metals	
Bottom Characteristics			
	*AVS *sediment contaminants (organics, metals)		

of seasonal differences in the benthic macroinvertebrate assemblages. Organisms are identified to genus and species. Water column and bottom characteristics are also measured to evaluate the status of physicochemical conditions.

10.1.1 Sampling Procedure

Primary objectives of Tier 3 benthic infaunal sampling are to evaluate potential impairment to this assemblage, to develop and refine biocriteria, to diagnose causes and sources of observed impairment, and to evaluate seasonal changes in the benthic infauna. This tier includes more frequent sampling (a minimum of four times per year) than either Tiers 1 or 2 to allow detailed discrimination of seasonality of benthic abundance. See Section 8.1.1 for full detail on sampling procedures.

10.1.2 Index Period

Benthic infaunal macroinvertebrates are sampled once or twice during an appropriate index period, the timing of which is driven by the goals of the Tier 3 assessment and regional considerations. At least two or three other sampling visits are made throughout the remaining portion of the year to capture more detailed seasonal differences in benthos than would be possible in a Tier 2 assessment. Data collected in a previous Tier 2 assessment, or historic benthic infaunal macroinvertebrate data, can be used to determine the timing and frequency of non-index period sampling.

10.1.3 Analysis

Organisms in each sample are identified to genus and species. If desired, and resources are available,

ash-free dry weight, at least to the family level, may be measured to determine the viability of biomass-based metrics to the overall assessment. Other metrics should be selected by the resource management agency as appropriate based on historic data, data collected and metrics used in preceding tiers, and regional considerations.

10.2 Fish

Tier 3 assessment of the fish assemblage is intended to allow evaluation of impairment, to develop and refine biocriteria, to diagnose causes and sources of impairment, and to evaluate seasonal differences in the assemblage. Fish sampling in this tier can include shallow-water, pelagic, and demersal fish communities (Carmichael et al. 1992, Eaton and Dinnell 1994, Guillen 1994).

10.2.1 Sampling Procedure

See Section 8.2.1 for full description of fish sampling procedures.

10.2.2 Sample Processing

See Section 8.2.2 for full description of fish sample processing.

10.2.3 Analysis

Based on the enumerated species list, metrics selected by the state can be calculated to evaluate potential impairment to the fish assemblage, to develop or refine biocriteria, to examine seasonal dynamics of the assemblage, or to diagnose sources and causes of impairment.

10.3 Macrophytes

Tier 3 assessment of macrophytes is intended to provide sufficient data to

assess impairment to the macrophyte assemblage, to develop or refine biocriteria, or to diagnose sources and causes of impairment.

10.3.1 Sampling Procedure

See Section 8.3.1 for full description of macrophyte sampling procedures.

10.3.2 Index Period

See Section 8.3.2 for full description of the macrophyte index period.

10.3.3 Analysis

Percent cover and area may be derived from analysis of aerial photographs. Taxonomic identification from the field trips will allow development of a species list. Stem counts made within quadrats along each sampling transect in addition to biomass determination will provide more detailed information on assemblage condition. Detailed pathology observations should be made; they can be used to evaluate potential causes of impairment.

10.4 Phytoplankton

10.4.1 Sampling Procedure

See Section 9.4.1 for a full description of the phytoplankton sampling procedure.

10.4.2 Index Period

Phytoplankton should be sampled at least once during an appropriate index period and a minimum of three other times per year to capture seasonal changes in the composition and abundance of the assemblage. Following review of data collected from historical data or through any of

the assessment tiers described here, the resource management agency may determine that a higher frequency of sampling is needed to characterize the phytoplankton assemblage based on its potential for rapid spatial and temporal variation.

10.4.3 Analysis

See Section 9.4.3 for a full description of phytoplankton analysis.

10.5 Epibenthos (Developmental)

As in Tier 2, even though epibenthos is currently under development as a biological indicator, it can still be useful in the Tier 3 assessment.

10.5.1 Sampling Procedure

See Section 9.5.1 for a full description of the epibenthos sampling procedure.

10.5.2 Index Period

See Section 9.5.2 for a full description of the epibenthos index period.

10.5.3 Analysis

See Section 9.5.3 for a full description of epibenthos analysis.

10.6 Zooplankton (Developmental)

Zooplankton are an important link between phytoplankton in estuaries and coastal marine waters and higher consumers. States might choose to include this developmental assemblage as part of a Tier 3 assessment.

10.6.1 Sampling Procedure

Three replicate vertical tows using a 118-µm mesh net, 30-cm in diameter should be made at each sampling location. The tow should be vertically integrated; that is, starting from 0.5-m from the bottom to the surface, with a retrieval rate of 0.5- to 1-ms^{-1}. Collected organisms should be anesthetized with carbonated water and preserved in 4% formalin. For long-term storage after fixing, specimens should be transferred to 70% ethanol. APHA (1992) describes procedures for concentrating the samples and preparing them for examination.

10.6.2 Index Period

Zooplankton should be sampled once during an appropriate index period and a minimum of three other times during the year to capture seasonal variation in taxonomic composition and abundance.

10.6.3 Analysis

Samples should be identified to the lowest practical taxonomic level, preferably genus and species. Subsampling may be required to achieve reasonable numbers of organisms for identification.

10.7 Paleoenvironmental Systems (Developmental)

Developmental assessment of paleoenvironmental systems is intended to provide site-specific reference by showing past conditions. Several groups of organisms leave remains in the bottom sediments. Some of the remains are resistant to decay and become a permanent biological record of life in the estuary.

By comparing past biota with present-day biota, past environmental conditions can be inferred. Several groups of organisms have been used: diatoms, foraminifera, and dinoflagellate cysts. Of these, diatom frustules and foraminifera have been used most often, and most successfully, to infer past conditions. A sample of the top 1- to 2-cm of sediment contains a representative sample of diatoms from the most recent 1- to 5-years. If the sediments remain undisturbed, then remains preserved in the sediments are integrators of estuarine history (Charles et al. 1994, Dixit et al. 1992). Because of the developmental nature of this indicator, states or agencies wishing to use paleoenvironmental reconstruction should contact one of the laboratories engaged in this research for further information. The methods described here are intended to give a brief overview of the field, but should not be used to plan a monitoring program.

10.7.1 Sampling Procedure

Cores are generally taken with standard gravity corers, such as the K-B, Phleger, or Piston. The chosen corer should retrieve a core deep enough to sample sediments from the earliest desired time period, with minimal edge disturbance. Core length thus depends on time period and sedimentation rate. Core samples are extruded from the corer and subsectioned immediately after collection. Sections 1-cm thick are removed from the core at intervals according to the time resolution desired. These sections are removed from the core using an apparatus described by Glew (1988) then are bagged, labeled and identified using a permanent ink pen. The bags prepared from a single core sample

are placed in a sealed container for storage and transport. Samples are kept at 4°C until shipment.

10.7.2 Sample Processing

Once in the laboratory, the sections are dried and weighed. Foraminifera and diatoms are processed so as to digest organic matter and preserve carbonate (foraminifera) or silica (diatoms), following the standard methods of Krom and Berner (1983) and EMAP (USEPA 1994e). An aliquot of frustules or tests is mounted for optical and/or scanning electron microscopy for identification. Dinoflagellate cysts are subjected to a standard pollen analysis involving the digestion of minerals in cold HCl, followed by warm HF (adapted from Barss and Williams 1973). They are processed on a 10 µm sieve. Samples for counting and identification are not random, but systematic.

Transects are taken on microscope slides, counting and identifying all target taxa encountered. A count of 300 or more is necessary for meaningful analysis of percentage data, but lower counts are still valid if results are reported on a concentration basis. In some depositional systems it is not feasible to count 300 dinoflagellate cysts, but the data is still informative.

Charcoal is seen in pollen analysis and dinoflagellate cyst preparations. The larger sieve size used for foraminifera would exclude most charcoal particles, thus make this material unsuitable for charcoal studies.

10.7.3 Analysis

Standard dating methods use either carbon-14, pollen, ^{137}Cs, or ^{210}Pb (Dixit 1992, Cooper and Brush 1991, Dale et

al. 1999, Alve 1991). Additional time points can be established from traces of known historic events (charcoal from large-scale fires, radioisotopes from atmospheric testing and the Chernobyl accident). Known responses of indicator taxa or biogeochemical indicators (e.g., biogenic silica) are used to infer past environmental conditions of an estuary. This allows for the assessment of current environmental conditions based on those of the past.

Quantitative paleoenvironmental reconstruction in estuaries requires the development of a data set that associates current conditions with current surficial diatom, dinoflagellate, or foraminifera assemblages. Present-day associations are used to infer past conditions based on fossil assemblages in deeper sediment layers. Quantitative prediction is usually done in two steps: development of predictive models (calibration or transfer functions), followed by use of the models to infer environmental variables from fossil assemblages (Charles and Smol 1994). Quantitative reconstruction has not yet been widely developed for estuaries.

Chapter 11

Index Development

11.1 Overview

Many methods have been developed to assess the condition of water resources from biological data, beginning with the saprobien system in the early 20th century to present-day development of biological markers. This chapter will discuss three methods for analyzing and assessing water body condition from assemblage and community-level biological information:

1. *Multimetric index.*

This is the basis of many indexes used in fresh waters: the Index of Biotic Integrity (IBI; Karr et al. 1986), the Invertebrate Community Index (ICI; Ohio EPA 1987); the Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, Second Edition (RBP; Barbour et al. 1999); and state indexes developed from these (e.g., Southerland and Stribling 1995). More recently, multimetric IBI - type indexes have been developed for estuarine assemblages (e.g. Cape Cod fish, Deegan et al. 1997; Chesapeake Bay macroinvertebrates, Weisberg et al. 1997; Carolinian Province macroinvertebrates, Hyland et al. 1998). The Chesapeake Bay development (Weisberg et al. 1997) will be used to illustrate the method.

2. *Discriminant model index.*

This is the basis of stream bioassessment in Maine (Davies et al. 1993), and of the estuarine invertebrate indexes developed by the EMAP-NC program in the Virginian and Louisianan provinces (Weisberg et al. 1993; Schimmel et al. 1994; Strobel et al. 1994; Summers et al. 1993, 1994; Engle et al. 1994). The EMAP-NC, Louisianian and Virginian Province examples will be used to illustrate the method.

3. *Index derived from multivariate ordination.*

Smith et al. (2000) and Allen and Smith (2000) have developed a pollution tolerance index for near-coastal sites of Southern California, using species composition of benthic macroinvertebrates and demersal fish. Other approaches using ordination have demonstrated differences in composition between reference and stressed sites (e.g., Warwick and Clarke 1991). The approach of Smith et al. uses ordination of species composition to develop a numeric index on a scale of 0-100, that can be used directly for biocriteria. The Smith et al. example will be used to illustrate the method.

Many other methods are possible, as well as permutations of the three methods above, all of which are beyond the scope of this document. These three were selected because:

▸ They use community and assemblage data;

▸ The methods are not restricted to any one assemblage. The examples all use benthic macroinvertebrates, but any other assemblage could also be used, such as fish,

phytoplankton, zooplankton or macrophytes;

▸ The examples used to illustrate the methods have been carried out over wide geographic areas with many sites, demonstrating the generality of the methods;

▸ The examples used to illustrate the methods are concise, the methods were fully documented, and have been carried to completion, that is, assessment of biological impairment and non-impairment.

All three of the methods use the same general approach: sites are assessed by comparing the assemblage of organisms found at a site to an expectation derived from observations of many relatively undisturbed reference sites. The expectations are modified by classifying the reference sites to account for natural variability, and each assessment site is classified using non-biological (physical, chemical, geographic) information. Finally, metrics (methods 1 and 2) or the species ordination (method 3) are tested for response to stressors by comparison of reference and known impaired sites. An example of the assessment process is summarized in Figure 11-1.

This chapter will first discuss methods of classification, with emphasis on those that have been successful in estuaries and coastal waters. The remainder of the chapter then discusses the three assessment methods. This chapter is not intended to be an instruction manual on using the different statistical methods; it is intended to show, with selected examples, techniques that have been used to develop biological indexes. Details of applications and methodology can be found in the cited documents and articles and in statistical textbooks and manuals (e.g., Ludwig and Reynolds 1988, Reckhow and Warren-Hicks 1996).

11.2 Classification and Characterization of Reference Condition

The objective of characterization is to finalize the classification of reference sites and to describe (characterize) each of the reference classes in terms of metrics, assemblage composition, and physical-chemical variables. As outlined in Chapter 4, classification may be a physical rule-based classification, or an analytical data interpretation where rules are derived from the data. The analytical approach requires a relatively large reference data set to derive the classes and rules, with many sites and both biological and physical-chemical data from each site.

The basic assumption of classification is that biogeography, physical habitat, and water quality largely determine attributes such as taxa richness, abundance, and species dominance in estuarine and coastal marine biological communities. In other words, if habitats are classified adequately, reference biological communities should correspond to the habitat classification.

Several statistical tools can assist in site classification, but there is no one set procedure. If the rule-based classification is based on well-developed prior knowledge and professional judgment, graphical analysis of metrics, followed by any necessary modifications and tests of the resultant classification, it is usually sufficient. If necessary, the classification is refined until an optimal classification emerges that satisfactorily accounts for variation in reference site biological data.

If a physical classification is not self-evident, it may be necessary to develop an alternative classification from the data using one or more of several

1. Waterbody Classification — The physical and habitat data along with biological data are used to group reference sites into homogeneous classes	Partitioning of Entire Water Resource → Class 1, Class 2, Class N
2. Metric Identification — Those metrics or attributes that are ecologically relevant to assemblage and zoogeography are identified	Identification of Biological Attributes → Metric 1 Value, Metric 2 Value, Metric N Value
3. Metric Calibration — Core metrics are sensitive to pollution and are informative of the ecological relationships of the assemblage to specific stressors or cumulative impacts	Evaluation and Calibration → Core Metric, Core Metric, Core Metric
4. Index Development — Core metrics, whose values vary in scale, are transformed to dimensionless numbers for aggregation	Aggregation → Index Score, Index Score
5. Threshold Establishment — The threshold (biocriterion) for discriminating between impaired and unimpaired is determined to provide a basis for assessment	Biocriteria Relative to Waterbody Class

Figure 11-1

The process for progressing from the classification of an estuary to assessing the health of the estuary. Adapted from Paulsen et al. 1991.

classification methods. These methods include cluster analysis and several ordination methods such as: principal components analysis, correspondence analysis, and multidimensional scaling.

11.2.1 Existing Classifications

With the growth of efforts to improve environmental monitoring and develop biocriteria, several successful classifications of estuarine and near-coastal biological assemblages have been developed. Here, we summarize several of these and integrate their findings on classification of North American estuarine assemblages.

EMAP Virginian Province

Natural environmental factors affecting species composition were examined in the EMAP Virginian Province Project (Paul et al. 1999, Strobel et al. 1995, Weisberg et al. 1993). Salinity has been known to control estuarine organisms since the early days of marine biology. Over 75% of the candidate measures were related significantly to salinity distributions (Figure 11-2). Correlation analysis was used to examine associations of habitat factors with candidate biological metrics. Of the correlations between candidate measures and habitat factors,

Figure 11-2

Mean number of species and salinity at EMAP-Estuaries sampling stations in the Virginian Province (from Weisberg et al. 1993). The regression line shown is the expected number of species based on the polynomial regression, and was used to estimate salinity-adjusted species richness measures.
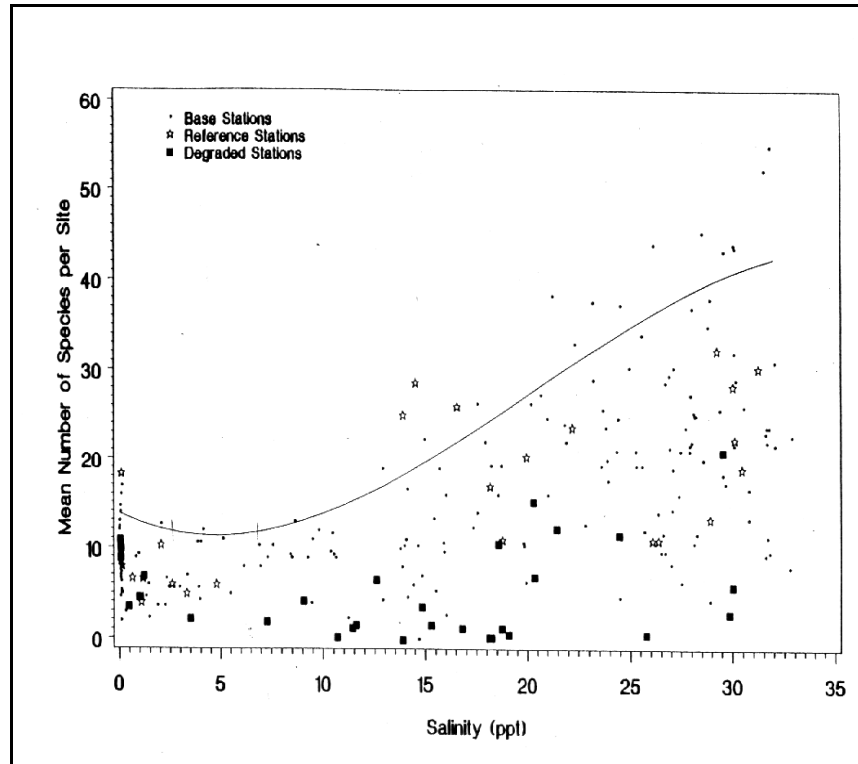
species richness was most strongly correlated with salinity.

In addition to salinity, the physical characteristics of estuarine sediments and depth also influence benthic infaunal distribution and the accumulation of contaminants in sediments (Rhoads 1974, Plumb 1981). EMAP collected sediment grain size, silt-clay content, latitude, and depth data to help interpret benthic response. Although silt-clay content and depth were statistically significant, only salinity was deemed to have a biologically significant influence on benthic macroinvertebrates (r^2 >0.025; Weisberg et al. 1993). Estuary type was stratified in the project design, but community differences due to estuary type were not reported.

Chesapeake Bay

Weisberg et al. (1997) developed an estuarine benthic index of biotic integrity for the Chesapeake Bay.

Cluster analysis of benthic infauna indicated seven distinct habitats defined by substrate and salinity. Polyhaline sand and mud, (salinities ($\geq$18 ‰) had the highest mean Shannon-Wiener diversities, at 4.0 and 3.55, respectively (Weisberg et al. 1997).

EMAP Carolinian Province

From July - September 1995, a study was conducted to assess the environmental condition of estuaries in the EMAP Carolinian Province (Hyland et al. 1998, see Chapter 13). The program sampled water depth, salinity, and substrate classifications (% silt-clay) as habitat indicators.

Species richness showed highly significant correlations with latitude, bottom salinity, and silt-clay/TOC sediment content. The Shannon-Wiener index, H′ (a combination of species richness and evenness), also showed highly significant correlations ($p \leq$ 0.0030) with bottom salinity as well as

with silt-clay fractions. As with diversity, infaunal abundance showed highly significant correlations ($p \leq 0.0016$) with the silt-clay and TOC sediment content (Hyland et al. 1998).

North Carolina

The North Carolina study was designed to compare biological metrics derived from three sampling methods (Ponar, epibenthic trawl, and sweep net). Salinity was the only habitat characteristic that was significantly correlated with biological metrics. Total taxa showed a positive correlation with salinity (Eaton 1994a; see Chapter 13).

Puget Sound

The objective of the Puget Sound study was to characterize benthic macroinvertebrate communities into habitats classified as degraded and habitats that are relatively unimpaired, which can then be classified as reference sites for the Sound (Llansó 1999).

The diverse assemblages sampled were mainly associated with sediment type and water depth, reinforcing results from previous studies (Lie 1974). The classes of sediments defined for the Puget Sound estuaries were: sands, clays, and mixed. These three classes did not have exact boundaries, but instead overlapped at both ends of their spectrums (Llansó 1999). Stations with finer substrates had fewer species than those with coarser substrates. On average, sand substrates supported more species and abundance than did clay, with deep sites having the lowest abundance levels. Overall, clay stations in the southern part of Puget Sound supported fewer species than many other shallow clay locations. The majority of species were not restricted to only one substrate, instead they were widely distributed in different types of sediment showing the most abundance in sand, mixed sediment, or muddy bottoms.

EMAP Louisianian Province

Prior studies in the Gulf of Mexico had shown salinity and sediment type to be among the most important factors that determine benthic infaunal relationships in Gulf of Mexico estuaries (Flint and Kalke 1985, Gaston et al. 1988, Rabalais 1990, Rakocinski et al. 1991). Of the 182 total sites sampled, Pearson correlations were performed between all candidate measures and salinity, longitude of sampling site (as a measure of geographical gradient), percent silt-clay, and total organic carbon content of sediments. Many of the correlations were statistically significant at $p < 0.05$, however, only salinity accounted for 20% or more of the variation (Summers et al. 1993).

Southern California Bight

The Southern California Coastal Water Research Project sampled megabenthic invertebrate assemblages, benthic infaunal assemblages, and demersal fish assemblages to determine their relationship to depth, latitude, and sediment types in the Southern California Bight. There was no salinity gradient because the entire study area was nearshore marine. Overall, depth was found to be the defining factor in the organization of each assemblage (Allen et al. 1999, Bergen et al. 1999).

Sediment type was found to be a secondary factor in the organization of benthic infaunal assemblages. This finding could be attributed to the large study area. In fact, within a constrained depth range, sediment type may be a more important factor (Bergen et al. 1999). These findings are consistent with those of Snelgrove and Butman

(1994) which suggested that the hydrodynamic environment and the amount of organic material in the sediment are more likely to be primary driving forces, with depth and sediment grain size as secondary correlates.

Conclusions

Three habitat indicators have been demonstrated repeatedly to influence biological assemblages of estuaries and near coastal environments. In studies where there was a salinity gradient, salinity was found to be the most important habitat indicator. Depth and substrate are also important and usually correlated, especially if there is a large depth gradient at the sample site. The physical type of estuary (e.g., fjord, lagoon, tidal river) has not been demonstrated to be vital in wide geographic studies, such as those conducted by EMAP in the Virginian, Louisianian, and Carolinian provinces, but may not have been adequately tested. Therefore, the importance of measuring estuary type, subregion, or subprovinces is still questionable.

Lessons learned from both EMAP and other independent studies conclude that the basic classification of an index should be by biogeograpical province, salinity, substrate (silt-clay content, sediment grain size), and depth. The effects of salinity, substrate, and depth should be tested within the study area to determine whether all are required as habitat indicators in an individual area. Moreover, decisions need to be made as to the use of discrete classes or continuous covariates in statistical analysis. If other classifications are suspected to be important indicators of the health of a system, they should also be tested (e.g., estuary type).

11.2.2 Assessing *a priori* Classifications

Although there is no serious doubt over the influence of salinity, sediment, and depth on estuarine biota, the effects must be characterized or calibrated to establish reference conditions. Several approaches have been used, as outlined in the examples in this chapter. Often, one of the first steps is a cluster analysis of the species composition of the sites to determine if sites can be broken down into groups (e.g., Weisberg et al. 1997, Smith et al. 2000). Sites may be divided into groups defined by the important variables (e.g., salinity and sediment; Weisberg et al. 1997, depth; Smith et al. 1999), or the groups may be separated by discriminant function analysis (DFA) if simple, single relationships are not sufficient (e.g., Engle and Summers 1999).

Another approach is to examine correlations between environmental variables and biological metrics calculated from the species data, so that reference expectations can be calibrated accordingly. For example, species richness in estuaries is strongly affected by salinity (refer to Figure 11-2). Weisberg et al. (1993) used the relationships of Figure 11-2 to develop a nonlinear regression of maximum expected species richness on salinity. Species richness was then adjusted by the salinity-specific maximum in further development of their model of impairment.

11.3 Index Development

An index for assessing sites can be developed after classification of sites of the region is completed. Index development using the three approaches followed in this chapter is discussed here.

11.3.1 Multimetric Index

Step 1. Identify Potential Measures For Each Assemblage.

Metrics allow the investigator to use meaningful indicator attributes in assessing the status of assemblages and communities in response to perturbation. The definition of a metric is a characteristic of the biota that changes in some predictable way with increased human influence (Barbour et al. 1995). For a metric to be useful, it must have the following technical attributes: (1) ecological relevance to the biological assemblage or community under study and to the specified program objectives, (2) sensitivity to stressors and provide a response that can be discriminated from natural variation. The purpose of using multiple metrics to assess biological condition is to aggregate and convey the information available regarding the elements and processes of aquatic communities.

All metrics that have ecological relevance to the assemblage under study and that respond to the targeted stressors are potential metrics for testing. From this "universe" of metrics, some will be eliminated because of insufficient data or because the range of values is not sufficient for discrimination between natural variability and anthropogenic effects. This step is taken to identify the candidate metrics that are most informative, and therefore, warrant further analysis.

Representative metrics should be selected from each of four primary categories: (1) richness measures for diversity or variety of the assemblage; (2) composition measures for identity and dominance; (3) tolerance measures that represent sensitivity to

perturbation; and (4) trophic or habit measures for information on feeding strategies and guilds. Table 11-1 further illustrates metrics for various assemblages that have been useful in estuaries. Components of Step 1 include:

▸ Review of the value ranges of potential metrics, and elimination of those that have too many zero values in the population of reference sites to calculate the metric at a large enough proportion of sites;

▸ Descriptive statistics (central tendency, range, distribution, outliers) to characterize metric performance within the population of reference sites of each site class;

▸ Elimination of metrics that have too high variability in the reference site population such that they cannot discriminate among sites of different condition.

Step 2. Select Robust Measures.

Core metrics are those that will discriminate between good and poor quality ecological conditions. Discriminatory ability of biological metrics is evaluated by comparing the distribution of each metric at a set of reference sites with the distribution of metrics from a set of "known" stressed sites (defined by physical and chemical characteristics) within each site class. If there is minimal or no overlap between the distributions, then the metric can be considered to be a strong discriminator between reference and impaired conditions (Figure 11-3).

Criteria are established to identify a population of "known" stressed sites based on physical and chemical measures of degradation. Criteria for

Table 11-1. Potential metrics for macrophytes, benthic macroinvertebrates, and fish that could be considered for estuaries. Redundancy can be evaluated during the calibration phase to eliminate overlapping metrics.

	Richness	**Composition**	**Tolerance**	**Trophic/Habitat**
Macrophytes	‣ Not applicable	‣ Not applicable	‣ TSS ‣ light attenuation ‣ Chlorophyll *a* ‣ DIN ‣ DIP	‣ % cover ‣ density of new shoots ‣ biomass ‣ stem counts
Benthic Macroinvertebrates	‣ dominant taxa ‣ taxa richness ‣ Shannon-Wiener Diversity Index ‣ mean # of species ‣ Pielou's Evenness Index	‣ # amphipods per event ‣ amphipod biomass ‣ mean abundance of bivalves/site ‣ # of gastropods per event	‣ % polychaetes ‣ polychaete biomass	‣ % or biomass epibenthic ‣ % or biomass deposit feeders ‣ % or biomass suspension feeders
Fish	‣ dominant taxa ‣ taxa richness ‣ # of estuarine spawners ‣ # anadronomous spawners ‣ total fish exclusive of Atlantic menhaden	‣ total # of species ‣ # species in bottom trawl ‣ # species comprising 90% of individuals	‣ #, % or biomass of menhaden	‣ Proportion of planktivores ‣ Proportion of benthic feeders ‣ Proportion of piscivores

stressed sites can include (Weisberg et al. 1993):

‣ Any sediment contaminant exceeding the Long et al. (1995) effects range-median (ER-M) concentration;

‣ Survival in toxicity tests less than 80% of controls;

‣ Low dissolved oxygen;

‣ Total sediment organic carbon > 3%.

Following identification of reference and stressed sites, the biological metrics that best discriminate between them are determined.

Those metrics having the strongest discriminatory power provide the most confidence in assessing biological condition of unknown sites. Metrics can be easily compared by estimating their discrimination efficiency (DE) or the percentage of stressed sites below a threshold representing the reference sites. For example, DE could be measured as the percentage of stressed sites below the 25th percentile of reference sites, for a given metric.

Several studies have used tests of statistical significance between reference and stressed sites to select metrics (e.g., Weisberg et al. 1997, Hyland et al. 1998). Significance tests should only be used if the sample size (number of reference and stressed sites) is large enough that the test has sufficient power to detect a meaningful difference.
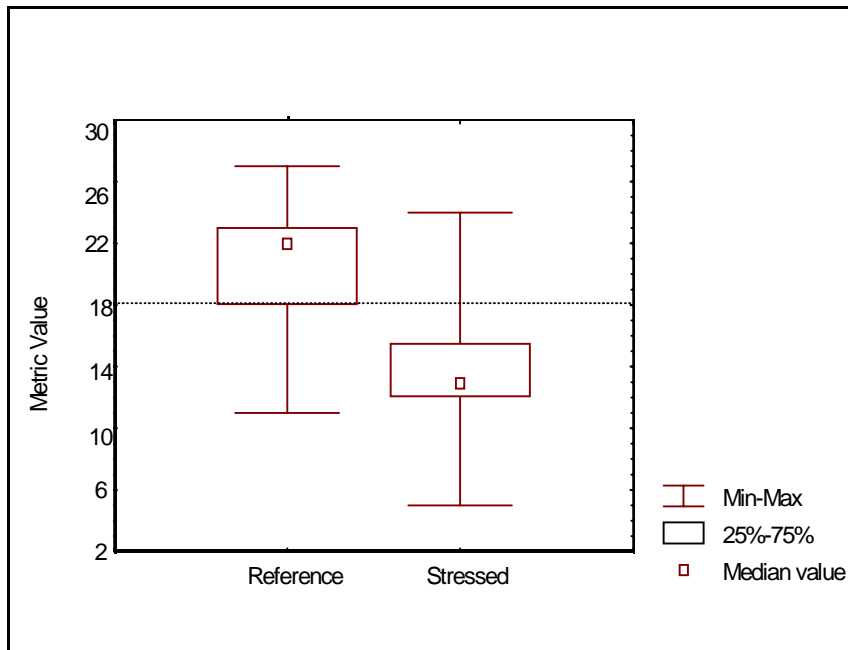
Figure 11-3

Hypothetical box plot illustrating how a successful metric discriminates between reference and stressed sites.

Step 3. Determine the best aggregation of core measures for indicating status and change in condition.

The purpose of an index is to provide a means of integrating information from the various measures of biological attributes (or metrics). Metrics vary in their scale—they are integers, percentages, or dimensionless numbers. Prior to developing an integrated index for assessing biological condition, it is necessary to standardize core metrics via transformation to unitless scores. The standardization assumes that each metric has the same value and importance; i.e., they are weighted the same, and that a 50% change in one metric is of equal value to assessment as a 50% change in another.

Where possible, the scoring criterion for each metric is based on the distribution of values in the population sites, which include reference sites; for example, the 95th percentile of the data distribution is commonly used to eliminate extreme outliers. From this upper percentile, the range of the metric values can be standardized as a percentage of the 95th

percentile value, or other (e.g., trisected or quadrisected), to provide a range of scores. Those values that are closest to the 95th percentile receive higher scores, and those having a greater deviation from this percentile receive lower scores. For those metrics whose values *increase* in response to perturbation the 5th percentile is used to remove outliers and to form a basis for scoring.

Alternative methods for scoring metrics are currently in use in various parts of the U.S. for multimetric indexes. A "trisection" of the scoring range has been well documented (Karr et al. 1986, Ohio EPA 1987, Weisberg et al. 1997, Hyland et al. 1998). More recent studies are finding that a standardization of all metrics as percentages of the 95th percentile value yields the most sensitive index, because more information of the component metrics is retained (e.g., Hughes et al. 1998).

Aggregation of metric scores simplifies management and decision making so that a single index value is used to determine whether action is needed. Biological condition of waterbodies is

judged based on the summed index value (Karr et al. 1986). If the index value is above a criterion, then the stream is judged as "optimal" or "excellent" in condition. The exact nature of the action needed (e.g., restoration, mitigation, pollution enforcement) is not determined by the index value, but by analyses of the component metrics in addition to the raw data, and integrated with other ecological information. Therefore, the index is not the sole determinant of impairment and diagnostics, but when used in concert with the component information, strengthens the assessment (Barbour et al. 1996b). Components of Step 3 include:

▸ Development of scoring criteria for each metric (within each site class) from the appropriate percentile of the data distribution (Figure 11-4). If the metric is associated with a significant covariate such as estuary size, depth, or salinity a scatterplot of the metric and covariate and a moving estimate of the appropriate percentile, are used to determine scoring criteria as a function of the covariate (e.g., Weisberg et al. 1993);

▸ Testing the ability of the final index to discriminate between populations of reference and anthropogenically affected (stressed) sites.

Step 4. Index thresholds for assessment and biocriteria.

The multimetric index value for a site is a summation of the scores of the metrics and has a finite range within each site class and index period, depending on the maximum possible score of the metrics (Barbour et al. 1996a). This range can be subdivided into any number of categories corresponding to various levels of impairment. Because the metrics are normalized to reference

conditions and expectations for the classes, any decision on subdivision should reflect the distribution of the scores for the reference sites.

Rating categories are used to assess the condition of both reference and non-reference sites. Most of the reference sites should be rated as *good* or *very good* in biological condition, which would be as expected. However, a few reference sites may be given the rating as *poor* sporadically among the collection dates. If a "reference" site consistently receives a fair or poor rating, then the site should be re-evaluated as to its proper assignment. Putative reference sites may be rated "poor" for several reasons:

▸ **Natural variability** — owing to seasonal, spatial, and random biological events, any reference site may score below the reference population 10th percentile. If due to natural variability, a low score should occur 10% of the time or less;

▸ **Impairment** — stressors that were not detected in previous sampling or surveys may occur at a "reference" site; for example, episodic non-point source pollution or historical contamination may be present at a site;

▸ **Non-representative site** — reference sites are intended to be representative of their class. If there are no anthropogenic stressors, yet a "reference" site consistently scores outside the range of the rest of the reference population, the site may be a special or unique case, or it may have been misclassified and actually belong to another class of sites.
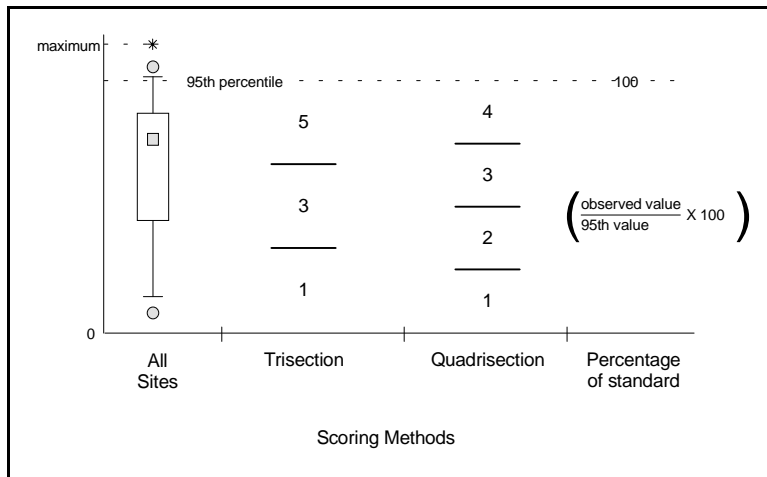
Figure 11-4

Basis of metric scores using the 95th percentile as a standard.

$$\left(\frac{\text{observed value}}{\text{95th value}} \times 100 \right)$$

Components of Step 4 include:

▸ Assessment categories are subdivided from the range of possible scores for each site class. Categories should be proportional to the interquartile range (or standard deviation) of total scores in the reference sites. Thus, reference sites with a small interquartile range (small s.d.; small coefficient of variation) would yield more assessment categories than a more variable reference population;

▸ The validity of biological condition categories is evaluated by comparing the index scores of the reference and known stressed sites to those categories. If reference sites are not rated good or very good, then some adjustment in either the biological condition designations or the listing of reference sites may be necessary;

▸ Confidence intervals are estimated for the multimetric index to help determine biological condition for sites that fall in close proximity to a threshold. Precision and sensitivity are determined from replicate samples, and are important for estimating the confidence of individual assessments.

Once the framework for bioassessment is in place, conducting bioassessments becomes relatively routine. Either a targeted design that focuses on site-specific problems or a probability-based design, which is appropriate for 305(b), area-wide, and watershed monitoring, can be done efficiently. Routine monitoring of reference sites should be based on a random selection procedure, which will allow for cost efficiencies in sampling while monitoring the status of the reference condition. Potential reference sites of each class would be randomly selected for sampling, so that an unbiased estimate of reference condition can be developed. A randomized subset of reference sites can be resampled at some regular interval (e.g., a 4-year cycle) to provide information on trends in reference sites.

Example 1: Chesapeake Bay Index Development

For the Chesapeake Bay, a separate benthic index was developed for each of seven habitats: tidal fresh, oligohaline, low mesohaline, high mesohaline sand, high mesohaline mud, polyhaline sand, and polyhaline mud (Weisberg et al. 1997). These habitats had been identified as separate assemblages in the classification step.

Reference and stressed sites were identified by the following: from existing Chesapeake Bay data, no reference sites could be in highly developed (urban) watersheds or near known point-source discharges, no reference site could have organic carbon content > 2%, no reference site could have any sediment contaminants exceeding the Long et al. (1995) effects range-median (ER-M) concentration, no reference site could have low dissolved oxygen, and no reference site could exhibit any sediment toxicity. Stressed sites were defined as those with any contaminant exceeding the ER-M concentration and measured sediment toxicity, or total organic carbon exceeded 3%, or dissolved oxygen was low, < 2-mgL^{-1} (Weisberg et al. 1997).

Index development proceeded through the steps:

Step 1. 17 candidate metrics were identified based on the paradigms of Pearson and Rosenberg (1978).

Step 2. 15 of the 17 metrics could distinguish between reference sites and stressed sites in one or more of the seven habitats.

Step 3. Four to seven of the metrics were used for an index specific to each habitat type. Scoring of metrics was on a 5-3-1 scale, with metric values greater than the reference site median scored as 5; between the 5th and the median of the reference sites scored as 3; and below the 5th percentile scored as 1.

Step 4. The index was able to correctly classify as reference or stressed 93% of an independent validation data set that had not been used to develop the index.

Example 2: Louisiana and Maryland Fish Indexes

Several states are developing fish indexes of biotic integrity (IBI) for estuarine species. The multimetric Index of Biotic Integrity (IBI) concept was originally developed for fresh water streams (Karr 1981), and has been modified and applied to a Louisiana estuary (Thompson and Fitzhugh 1986). The strength of this index is that many factors affecting biological integrity can be measured in fish (e.g., community composition, relative abundance, health, etc.). This proposed estuarine IBI maintains the same three main categories as those of the fresh water IBI: species composition, trophic composition, and fish condition. However, the metrics are modified to reflect estuarine habitats and fish assemblages. In addition, because estuarine systems exhibit a high degree of seasonality in their fish fauna, a measure of seasonal variability was incorporated. The metrics for estuaries are based on life history and habitat requirements similar to those of the fresh water IBI. Proposed metrics from Thompson and Fitzhugh (1986) for estuarine communities are listed in Table 11-2. A similar fish Index of Biotic Integrity is being adapted for application in estuarine and coastal marine habitats on the Gulf Coast of Texas (Guillen 1995).

The state of Maryland has also developed a fish Index of Biotic Integrity that is more rapid and less expensive to apply (Jordan et al. 1992). This fish IBI is comprised of nine metrics (Table 11-3) that can be compared to measurements of the physical environment such as dissolved oxygen and land use.

Table 11-2.   Estuarine fish IBI metrics proposed by Thompson and Fitzhugh (1986).

Community Structure/Function	Metric
Species composition	Total number of fish species Number and identity of resident estuarine species Number and identity of marine species Number and identity of sciaenids Number and identity of freshwater species Proportion of individuals as bay anchovy Measure of seasonal overlap of fish community Number of species needed to make up 90% of collection
Trophic composition (for adults of species)	Proportion of individuals as generalized benthic feeders Proportion of individuals as generalized plankton grazers Proportion of individuals as top carnivores
Fish abundance and condition	Proportion of young of year in sample or number of individuals in sample Proportion of individuals with disease, tumors, fin damage, and other anomalies

The results of some preliminary analyses from areas in the Chesapeake Bay with salinities ranging from 0-to-16 ppt indicate that the Maryland IBI can be used to identify large scale spatial and temporal trends in biological integrity and that the index responds to water quality (DO) and land use impacts.

11.3.2 Discriminant Model Index

Discriminant Model Approach

The discriminant model approach was used by EMAP to develop benthic condition indexes for the Virginian Province (Mid-Atlantic) and for the Louisianian Province (Gulf Coast) (Engle et al. 1994, Summers et al. 1993, Weisberg et al. 1993, Paul et al. 1999) based on defined reference sites. Sets of minimally impaired sites; i.e., "reference" and impaired sites were identified; impaired sites were affected by either hypoxia (DO <2 mgL^{-1}); toxic sediments; or sediment contamination above the ER-M threshold. Minimally impaired sites were defined to have DO >5 mgL^{-1} and no detectable toxicity or contamination. The two site types represented the ends of a continuum,

with intermediate sites not used for discriminant model building (Engle et al. 1994, Weisberg et al. 1993).

The classification step for the EMAP discriminant models consisted of examining associations between benthic macroinvertebrate metrics and physical habitat measures of salinity, sediment grain size, and depth. Only salinity had a strong relationship with the taxa richness metric; taxa richness was estimated as the percent of taxa expected, adjusted for salinity (refer to Figure 11-2).

Discriminant Model Analysis

The discriminant model analysis is a multivariate procedure that attempts to build a model that will predict the membership of a site into two or more predetermined classes. In the example used in EMAP, the classes were reference and impaired sites (by low DO, toxicity or metal contamination). The model procedure attempts to find a linear combination of input variables (biological metrics) that best predicts membership in the class. Alternative

---

Estuarine and Coastal Marine Waters: Bioassessment and Biocriteria Technical Guidance   11-13

Table 11-3. Maryland estuarine fish IBI metrics.

Community Structure/Function	Metric	Response to Impairment
Species composition	Total number of species Number of species in bottom trawl Number of species comprising 90 percent of individuals	reduced reduced reduced
Trophic composition (for adults of species)	Proportion of planktivores Proportion of benthic feeders Proportion of piscivores	increased decreased decreased
Fish abundance and condition	Number of estuarine spawners Number of anadromous spawners Total fish exclusive of Atlantic menhaden	decreased decreased decreased

models are tested by estimating the proportion of sites (from the model-building data set) that are misclassified. The best model usually has the lowest misclassification rate. A test of a model requires an independent test data set that was not used to build the model.

EMAP built discriminant models using benthic metrics in a stepwise model building approach. The models used three to five metrics in the Louisianian and Virginian provinces respectively, and both models used taxa richness (Engle et al. 1994, Weisberg et al. 1993). The benthic indexes were then calculated as the discriminant score of a site and standardized on a scale of 1 to 10.

Performance of the discriminant models was good in distinguishing reference from impaired sites in the calibration data: 100% for the Gulf of Mexico sites (Engle et al. 1994; n = 16 sites) and 86-93% for the Virginian Province sites (Weisberg et al. 1993; n = 33 sites). When tested with validation data collected in subsequent years, however, both sets or models failed to predict adequately and had to be redeveloped (Engle and Summers 1999, Strobel et al.

1994, 1995, Paul et al. 1999). Inclusion of several years of monitoring data in both provinces produced more robust and reliable models. In the Virginian Province, the robust calibration data set consisted of 60 sites (30 each).

An improved index was created to be applicable across a variety of estuarine environments in the Gulf of Mexico (Engle and Summers 1999). The statistical approach described in Engle and Summers (1999) proved to be applicable throughout the estuaries in the northern Gulf of Mexico. This benthic index was also validated independently by Rakoncinski (1997), who compared results of canonical correspondence analysis (CCA) with data from EMAP-E (1991-1992), using the index developed in Engle et al. 1994 (Engle and Summers 1999).

11.3.3 Index Derived from Multivariate Ordination

An index for biocriteria was derived by Smith et al. (2000) using multivariate ordination to derive a pollution gradient, which in turn was used to develop an index. The approach was developed with benthic macroinvertebrates from the Southern

California Bight (Smith et al. 2000; see also 11.2.1, p. 11-5), and is currently being applied to demersal fish from the same waters (Allen and Smith 2000). The approach is computationally intensive and rather complex. We will describe the result first (the index and its components), and then briefly describe how the components themselves are derived.

The central assumption of this approach is that each species has a tolerance for pollution, and that if the pollution tolerance is known for sufficiently large set of species, it is possible to infer the degree of degradation from species composition and the tolerances. This is the basis of the familiar Hilsenhoff Biotic Index (HBI; Hilsenhoff 1987) of freshwater bioassessment, as well as of several metrics in the multimetric approach. For example, capitellid polychaetes are known to be tolerant to organic pollution (BOD). The index used by Smith et al. is a weighted average tolerance value of all species found in a sample, weighted by abundance of the species:

Equation 11-1.

$$I_s = \frac{\sum_{i=1}^{n} a_{si}^{f} \, p_i}{\sum_{i=1}^{n} a_{si}^{f}}$$

where I_s is the index value for sample s, n is the number of species in sample s, a_{si} is the abundance of species i in sample s, p_i is the tolerance value of species i, and the exponent f is used to downweight extreme abundances. If f is zero, then the index is not weighted by abundance (Smith et al. 2000, Allen and Smith 2000).

The index of equation (11-1) is computationally almost identical (except for the introduction of the transformation exponent f) to the Hilsenhoff Biotic Index. Biocriteria can be assigned to index values; for example, if the index is defined in the range from 0 (unpolluted) to 100 (severely polluted), then a criterion for Class A estuarine waters might be values ≤ 25.

The steps below outline the derivation of the tolerance values p_i. A data set is required with sites that span a range from unpolluted to severely polluted. In the Southern California Bight, these were defined by sediment contaminant levels above and below the effects range median (ER-M) and effects range low (ER-L) concentrations (Long et al. 1995). Levels of a contaminant below the ER-L, between the ER-L and ER-M, and above the ER-M are rarely, occasionally, and frequently, respectively, associated with adverse effects. Impacted sites had six of eight selected contaminants (Cu, Pb, Ni, Zn, Cd, Cr, PCB, and DDT) above the ER-M. Reference sites consisted of stations lying outside of POTW discharge areas and with no more than one selected contaminant above the ER-L for a contaminant.

The data must be divided into two sets: a calibration subset and a test subset.

Step 1. Ordination analysis of species abundance (calibration data).

Ordination analysis produces a plot of sites in ordination space (Figure 11-5). Distances between pairs of points are proportional to the dissimilarity of species composition in the corresponding samples: samples with very similar composition will be close together in the ordination diagram. If the species are associated with the pollution gradient, the sites will define a gradient, with polluted sites at one end
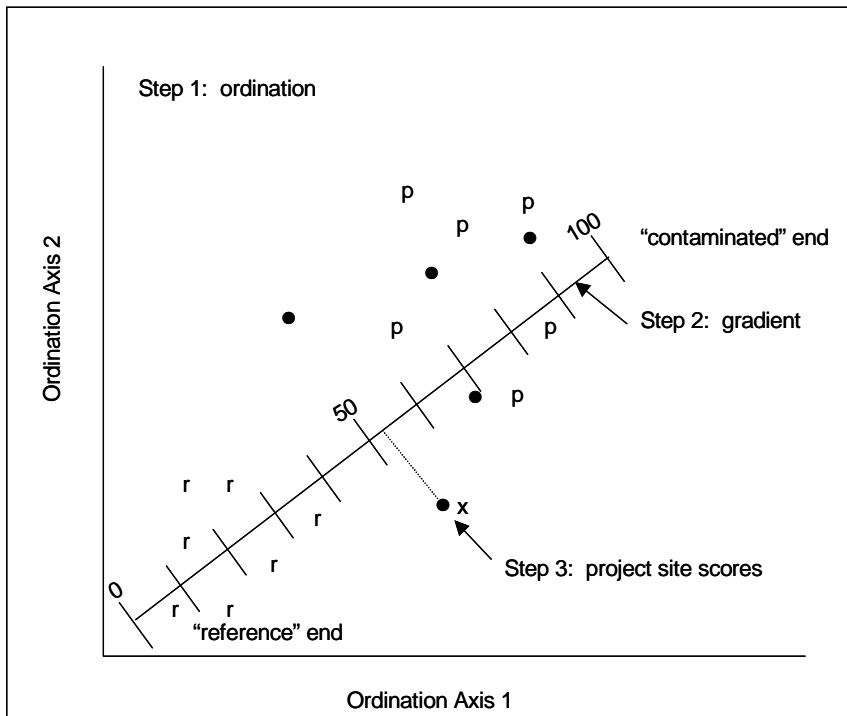
Figure 11-5

Steps 1-3. Establishing site scores on a contamination gradient. The gradient is established between reference ("r") and contaminated ("p") sites as plotted in ordination space. Dots are sites not designated as either reference or contaminated. The projection of site "x" on the gradient (dotted line) yields its site score. Adapted from Smith et al. 2000.

and unpolluted sites at the other (Figure 11-5).

Step 2. Find the pollution gradient.

The two ends of the pollution gradient are defined as the average positions in ordination space of the unpolluted and polluted sites, respectively. These ends are connected by a line, which represents the pollution gradient as expressed by the observed species compositions.

Step 3. Project all calibration observations onto the pollution-effects gradient.

The position of each site in the ordination space is projected onto the gradient. This projection is the site score of the calibration sites (Figure 11-5).

Step 4. Rescale the projections.

Site scores are scaled from 0 ("least polluted") to 100 ("most polluted").

Step 5. Compute tolerance values for each species.

Each species has an "average" position on the pollution gradient. These species positions are the tolerance values (p_i) of Equation 11-1. Site scores calculated in Step 4 give each site a position on the pollution gradient. The abundance of a species at each observation can be plotted against the site scores (Figure 11-6). The species position on the gradient, or the tolerance value p_i, is the abundance-weighted average position for the species over all sites.

Step 6. Compute the f parameter.

The f-parameter is iterated simultaneously with the p_i in an optimization procedure (Smith et al. 2000).

The species tolerance scores were in turn used to predict the Benthic Response Index (BRI) according to Equation 11-1. The BRI is the position of a site on the contamination gradient, or the predicted value of the site score
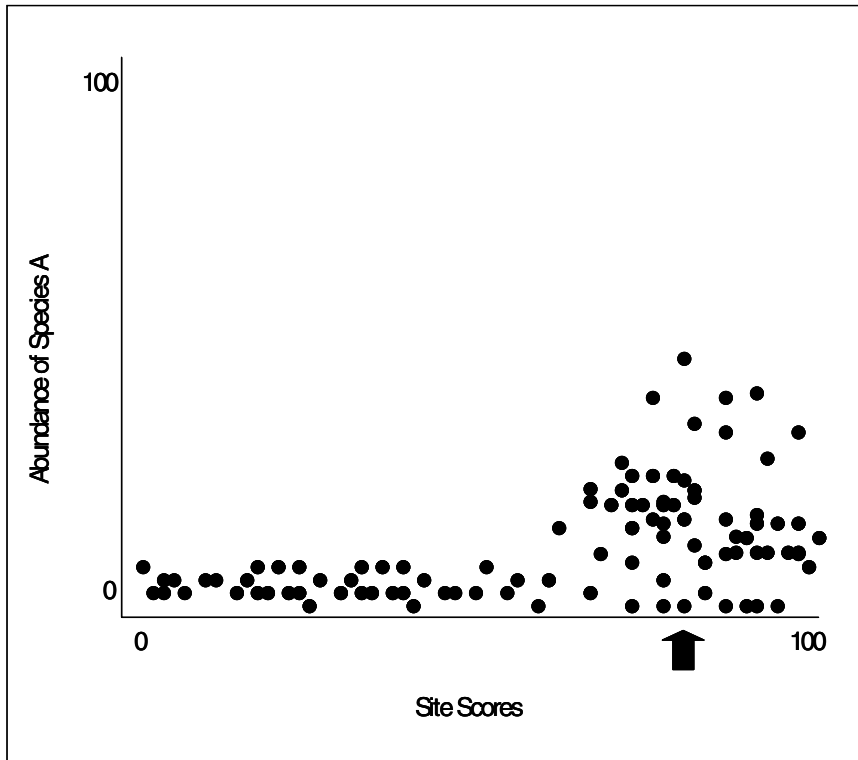
Figure 11-6

Step 5. Computing the tolerance values. Abundance of Species A and site scores (from Figure 11-5) of all sites where Species A occurs. The abundance-weighted average score over all sites is Species A's pollution tolerance score (arrow). This example shows a highly tolerant species, which occurs in greatest abundances at the most polluted sites. Adapted from Smith et al. 2000.

calculated in Step 4. Actual site scores (Step 4) are calculated only for the calibration data; site score is predicted as BRI for all assessment sites.

The BRI was developed separately for 3 depth zones: 10-35-m, 25-130-m, and 110-324-m. Earlier work had shown that benthic communities off Southern California could be classified by depth and sediment type (see Section 11.2.1, p. 11-5). Sediment type was secondary, and was not deemed to have a strong enough effect to justify further categorization of the data set.

The tolerance index developed by Smith et al. was then tested with the independent data (not used to develop the index). The independent test showed that the model was largely correct in predicting position along the contamination gradient. For further details of calculations and formulas, see Smith et al. (2000). The approach is currently being extended to demersal fish from the same region (Allen and Smith 2000).

Smith et al. estimated tolerance values for over 450 marine species from southern California. The BRI contamination score can be calculated for any new site from species abundance data at the site. The BRI has a range from 0 (unpolluted) to 100 (severely polluted) and biocriteria can be set at selected values for specific aquatic life uses of coastal waters of Southern California. Reference sites had BRI values < 25, and all severely contaminated sites had BRI > 36 (Smith et al. 2000). The reference values could form the basis of biocriteria for the region.

Chapter 12

Quality Assurance: Design, Precision and Management

Quality assurance (QA) is an integrated program for ensuring the reliability of monitoring and measurement data and includes quality control. Quality control (QC) refers to operational procedures for obtaining prescribed standards of performance in the monitoring and measurement process. Specific QC elements can be developed for most, if not all, project activities. All project activities, from sampling (data collection) and laboratory analysis to statistical analysis and reporting, are potential error sources (Peters 1988). Because error is cumulative and can significantly affect the results of a project, all possible efforts must be made to control it. Therefore, quality assurance is a continuous process that should be implemented throughout the entire development and operation of a program.

The purpose of an overall quality assurance project plan (QAPP), containing specific QC elements and activities, is to minimize—and when possible eliminate—the potential for error. Additionally, there are objective mechanisms for evaluating activities relative to pre-established measurement quality objectives and other project goals. The appropriateness of the investigator's methods and procedures and the quality of the data to be obtained must be ensured before the results can be accepted and used in decision making.

QA is accomplished through:

▸ Program design;

▸ Investigator training;

▸ Standardized data gathering and processing procedures;

▸ Verification of data reproducibility;

▸ Instrument calibration and maintenance.

As outlined below, QA requirements apply to all activities in an ecological study. More detailed guidance and examples for QA activities should be obtained from USEPA (1994c, and 1998a); more general guidance is outlined by USEPA (1993b).

12.1 Program Design

A central component of QA is overall study design which includes formulation of questions and hypotheses, experimental design, and development of analysis approaches. The classical approach by which scientists plan research consists of the following steps:

▸ Statement of the problem to be resolved;

▸ Formulation of alternative hypotheses that will explain the phenomena or, in the case of

problems that do not involve elaboration of processes, formulation of specific research questions;

▸ Establishment of boundaries within which to resolve the problem;

▸ Formulation of an experimental or study design that will falsify one or more hypotheses or answer the specific research questions;

▸ Establishment of uncertainty limits including setting acceptable probabilities of type I and type II errors for statistical hypothesis testing;

▸ Optimization of the study design including power analysis of the statistical design.

Experimental advances in basic sciences have not included the last two steps because uncertainty limits were inappropriate or unknown. Examination of experimental advances also reveals that a high degree of creativity and insight is required to formulate hypotheses and study designs; no formal planning process or "cookbook" can guarantee creativity and insight. Nevertheless, documentation of the planning process and a complete explanation of the conceptual framework help others evaluate the validity of scientific and technical achievements.

12.1.1 Formulation of a Study Design

A study design is developed to answer the specific monitoring questions developed in formulating the questions and objectives. Sampling design considerations were discussed in Chapter 5.

For quality assurance, some effort will always be required for repeated samples so that measurement error can always be estimated from a subset of sites. Repeated measurement at 10% or more of sites is common among many monitoring programs.

12.1.2 Establishment of Uncertainty Limits

The level of uncertainty associated with environmental measurements (due to natural variability, sampling error, measurement error, or other sources of uncertainty) propagates directly to the uncertainty of inferences and conclusions that can be made from the data. Establishing the limits of statistical uncertainty for conclusions also sets limits for the data to be collected (also known as Data Quality Objectives [DQOs]; Chaloud and Peck 1994). As mentioned in Chapter 5, there is a close association between sampling intensity and uncertainty. Reducing uncertainty usually results in greater costs. Assessing uncertainty, and optimizing the study design (below) require at least pilot data in hand, if not results from one year or more of monitoring.

As an example of uncertainty limits, USEPA's EMAP program established the following (Chaloud and Peck 1994):

▸ Estimate the status of a population of resources with 95% confidence intervals that are within 10% of the estimate;

▸ Determine average change in status of 20% over 10 years with 95% confidence and statistical power of 0.8.

EMAP selected 95% confidence intervals, however, there is nothing "scientific" about choosing 95% intervals over, say, 90% or 99%. The second limit above, determining

change, implies that EMAP managers were only willing to conclude a false change in status 1 time out of 20 (Type I error; false positive), but were willing to conclude a false lack of change 1 time out of 5 (Type II error, false negative).

12.1.3 Optimizing the Study Design: Evaluation of Statistical Power

A principal aspect of probability sampling is determining how many samples will be required to achieve the monitoring goals and what is the probability of making an incorrect decision based on the monitoring results. The primary tool for conducting these analyses is statistical power analysis. Evaluating statistical power is key to developing data quality criteria and performance specifications for decision making (USEPA 1996b) as well as evaluating the performance of existing monitoring programs (USEPA 1992). Power analysis provides an evaluation of the ability to detect statistically significant differences in a measured monitoring variable. The importance of this analysis can be seen by examining the possible outcomes of a statistical test. The null hypothesis (H_o) is the root of hypothesis testing. Traditionally, null hypotheses are statements of no change, no effect, or no difference. For example, the mean abundance at a test site is equal to the mean abundance of the reference sites. The alternative hypothesis (H_a) is counter to H_o, traditionally being statements of change, effect, or difference. Upon rejecting H_o, H_a would be accepted.

The two types of decision errors that could be made in hypothesis testing are depicted in Table 12-1. A Type I error (i.e., false positive) occurs when H_o is rejected although H_o is really true. A Type II error (i.e., false negative) occurs when H_o is not rejected although H_o is

really false. The magnitude of a Type I error is represented by α and the magnitude of a Type II error is represented by β. Decision errors are the result of measurement and sampling design errors that were described in Section 12.1.1. A proper balance between sampling and measurement errors should be maintained because accuracy limits effective sample size and vice versa (Blalock 1979).

Comparison of Significance Level and Power

Regardless of the statistical test chosen for analyzing the data, the analyst must select the significance level of the test. That is, the analyst must determine what error level is acceptable. The probability of making a Type I error is equal to the significance level (α) of the test and is selected by the data analyst. In many cases, managers or analysts define 1-α to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99%), although there have been environmental applications where 1-α has been set to 0.80. Selecting a 95% confidence level implies that the analyst will reject the H_o when H_o is really true, i.e., a false positive, 5% of the time.

Type II error depends on the significance level, sample size, number of replicates, variability, and which alternative hypothesis is true. The power of a test (1-β) is defined as the probability of correctly rejecting H_o when H_o is false. In general, for a fixed sample size, α and β vary inversely. Power can be increased (β can be reduced) by increasing the sample size or number of replicates. Figure 12-1 illustrates this relationship. Suppose the interest is in testing whether there is a significant difference between the means from two independent random samples. As the difference in the two sample means increases (as indicated on

Table 12-1. Errors in hypothesis testing.

Decision	State of the population (truth)	
	H_o is True	H_o is False
Accept H_o	$1-\alpha$ (Confidence level)	β (Type II error)
Reject H_o	α (Significance level) (Type I error)	$1-\beta$ (Power)

Figure 12-1

Effect of increasing sample size from n_1 to n_2 on power. The curves represent the probability distribution of the sample means from 2 samples, reference and test, and for 2 sample sizes n_1 and n_2 where $n_2 > n_1$.
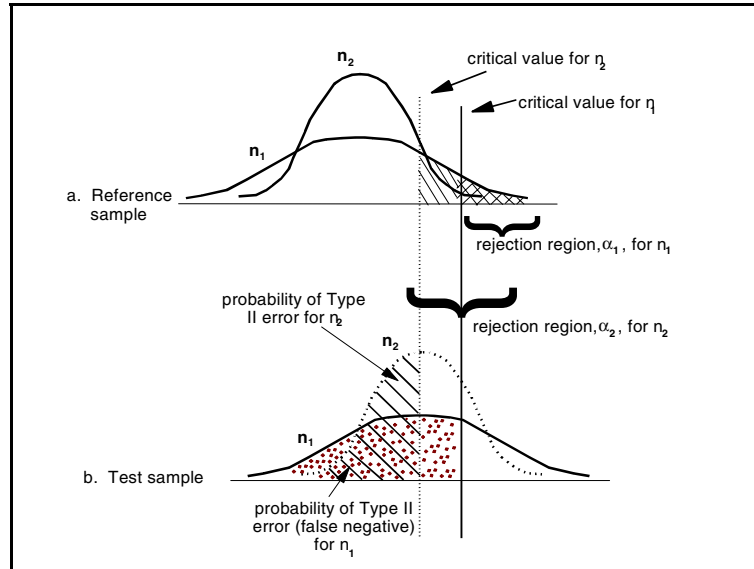
the *x*-axis), the probability of rejecting H_o, the power, increases. If the real difference between the two sample means is zero, the probability of rejecting H_o is equal to the significance level, α. Figure 12-1a shows the general relationship between α and β if α is changed. Figure 12-1b shows the relationship between α and β if the sample size is increased. The tradition of 95% confidence ($\alpha = 0.05$) is entirely arbitrary; there is no scientific requirement that confidence be set at 95%. Indeed, for environmental protection, power is at least as important—and possibly more important—than confidence (Peterman 1990, Fairweather 1991).

Basic Assumptions

Usually, several assumptions regarding data distribution and variability must be made to determine the sample size. Applying any of the equations described in this chapter is difficult when no historical data set exists to quantify initial estimates of proportions, standard deviations, means, or coefficients of variation. To estimate these parameters, Cochran (1963) recommends four sources:

▸ Existing information on the same population or a similar population;

▸ A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of

the second step. Use data from both steps to estimate the final precision of the characteristic(s) sampled;

▸ A "pilot study" on a "convenient" or "meaningful" subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because often the pilot sample is not representative of the entire population to be sampled;

▸ Informed judgment, or an educated guess.

For evaluating existing programs, proportions, standard deviations, means, etc. would be estimated from actual data.

Some assumptions might result in sample size estimates that are too high or too low. Depending on the sampling cost and cost for not sampling enough data, it must be decided whether to make conservative or "best-value" assumptions. Because of the fixed mobilization costs, it is probably cheaper to collect a few extra samples the first time than to realize later that additional data are needed. In most cases, the analyst should probably consider evaluating a range of assumptions regarding the impact of sample size and overall program cost. USEPA recommends that if the analyst lacks a background in statistics, he/she should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

Simple Comparison of Proportions and Means from Two Samples

The proportion (e.g., percent dominant taxon) or mean (e.g., mean number of EPT taxa) of two data sets data sets can be compared with a number of statistical tests including the parametric two-sample t-test, the nonparametric Mann-Whitney test, and two-sample test for proportions (USEPA 1996b). In this case, two independent random samples are taken and a hypothesis test is used to determine whether there has been a significant change. To compute sample sizes for comparing two proportions, p_1 and p_2, it is necessary to provide a best estimate for p_1 and p_2, as well as specifying the significance level and power ($1\text{-}\beta$). Recall that power is equal to the probability of rejecting H_o when H_o is false. Given this information, the analyst substitutes these values into the following equation (Snedecor and Cochran 1980):

Equation 12-1.

$$n_o = (Z_\alpha + Z_{2\beta})^2 \, \frac{(p_1 q_1 + p_2 q_2)}{(p_2 - p_1)^2}$$

where Z_α and $Z_{2\beta}$ correspond to the normal deviate. Common values of $(Z_\alpha + Z_{2\beta})^2$ are summarized in Table 12-2. To account for p_1 and p_2 being estimated, t could be substituted for Z. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 12-1; (2) round n_o up to the next highest integer, f; and (3) multiply n_o by $(f+3)/(f+1)$ to derive the final estimate of n.

To compare the mean from two random samples to detect a change of δ; i.e., $\bar{x}_2\text{-}\bar{x}_1$, the following equation is used:

Equation 12-2.

$$n_o = (Z_\alpha + Z_{2\beta})^2 \, \frac{(s_1^2 + s_2^2)}{\delta^2}$$

Common values of $(Z_\alpha + Z_{2\beta})^2$ are summarized in Table 12-2. To account

Table 12-2. Common values of $(Z_\alpha + Z_{2\beta})^2$ for estimating sample size for use with Equations 12-1 and 12-2 (Snedecor and Cochran 1980).

Power, 1-β	α for One-sided Test			α for Two-sided Test		
	0.01	0.05	0.10	0.01	0.05	0.10
0.80	10.04	6.18	4.51	11.68	7.85	6.18
0.85	11.31	7.19	5.37	13.05	8.98	7.19
0.90	13.02	8.56	6.57	14.88	10.51	8.56
0.95	15.77	10.82	8.56	17.81	12.99	10.82
0.99	21.65	15.77	13.02	24.03	18.37	15.77

for s_1 and s_2 being estimated, Z should be replaced with t. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute n_o using Equation 12-2; (2) round n_o up to the next highest integer, f; and (3) multiply n_o by $(f+3)/(f+1)$ to derive the final estimate of n.

A special case of Equation 12-2 arises for biocriteria, when we compare the mean of a sample to determine if the value is below some set limit, that is, if the site is impaired or below a reference threshold. The threshold is fixed by previous investigations and decisions, and is not a random variable. We ask now whether we can detect a change of δ; i.e., $C-\bar{x}_1$, where C is the biocriteria limit:

Equation 12-3.

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(s_1^2)}{\delta^2}$$

In Equation 12-3, Z_α is most often one-tailed, because the concern is only whether the value is below the threshold.

Sample Size Calculations for Means and Proportions

For large sample sizes or samples that are normally distributed, symmetric confidence intervals for the mean are appropriate. This is because the distribution of the sample mean will approach a normal distribution even if the data from which the mean is

estimated are not normally distributed. The Student's t statistic ($t_{\alpha/2,n-1}$) is used to compute symmetric confidence intervals for the population mean, μ:

Equation 12-4.

$$\bar{x} - t_{\alpha/2,n-1} \sqrt{s^2/n} \le \mu \le \bar{x} + t_{\alpha/2,n-1} \sqrt{s^2/n}$$

This equation is appropriate if the samples are normally distributed or the sample size is greater than 30 (Wonnacott and Wonnacott 1969), although Helsel and Hirsch (1992) suggest that highly skewed data might require more than 100 observations.

Although several approaches exist to estimate confidence levels for any percentile, many rely on assuming a normal or lognormal distribution. The approach presented here (Conover 1980) for more than 20 observations does not rely on these assumptions. Conover (1980) also provides a procedure for smaller sample sizes. To calculate the confidence interval corresponding to the median, lower quartile, or upper quartile, the following procedure is used.

1. Order the data from smallest to largest observation such that

$$x_1 \le \ldots \le x_r \le \ldots \le x_p \le \ldots x_s \le \ldots \le x_n$$

Quality Assurance: Design, Precision and Management

where x_p corresponds to the median; i.e., p=0.5, lower quartile; i.e., p=0.25, or upper quartile; i.e., p=0.75.

2. Compute the values of r^* and s^* as

   Equation 12-5.
   $$r^* = np + Z_{a/2} \ (np(1-p))^{0.5}$$
   $$s^* = np + Z_{a/2} \ (np(1-p))^{0.5}$$

   where $Z_{\alpha/2}$ is selected from a normal distribution table.

3. Round r^* and s^* up to the next highest integers r and s. The 1-α lower and upper confidence limits for x_p are x_r and x_s, respectively.

It can be seen from Equation 12-5 that estimation of medians or quartiles from small samples can result in large confidence intervals for the estimate. For example, the 90% confidence interval for the lower quartile of a sample of n=10 covers the first 5 observations. A sample of less than 10 observations would have a confidence interval extending below the smallest observation. This is the reasoning behind a general "rule of thumb" that estimation of reference conditions should be based on a sample of 10 or more sites, if at all possible. Figure 12-2 gives example sample size calculations for comparing proportions and population means.

12.2 Management

12.2.1 Personnel

Trained and experienced biologists should be available to provide thorough evaluations, provide support for various activities, and serve as QC checks. They should have training and experience commensurate with the needs of the program. At least one staff member should be familiar with establishing a QA framework. QA programs should

document personnel responsibilities and duties and clearly delineate project organization and lines of communication (USEPA 1998a). A time line illustrating completion dates for major project milestones or other tasks can be a tremendously useful tool to track project organization and progress.

12.2.2 Resources

Laboratory facilities, adequate field equipment, supplies, and services should be in place and operationally consistent with the designed purposes of the program so that high-quality environmental data can be generated and processed in an efficient and cost-effective manner (USEPA 1992). Adequate taxonomic references and scientific literature should be available to support laboratory work, data processing, and interpretation.

12.3 Operational Quality Control

Protocols should be developed for designing a data base and for screening, archiving, and documenting data. Data screening identifies questionable data based on expected values and obvious outliers. Screening is especially important if data are gathered from a variety of sources and the original investigators and data sheets are no longer available. Figure 12-3 defines the qualitative and quantitative data characteristics that are most often used to describe data quality.

These measurement quality indicators require *a priori* consideration and definition before the data collection begins. Taken collectively, they provide a summary characterization of the data quality needed for a particular environmental decision. Duplication of approximately 10% of the total sampling effort is a common level for

Figure 12-2

Example
sample size
calculations for
comparing
proportions
and population
means.

Example 1—Sample size calculation for comparing proportions

To detect a difference in proportions of 0.20 with a two-sided test, α equal to 0.05, 1-β equal to 0.90, and an estimate of p_1 and p_2 equal to 0.4 and 0.6, n_o is computed from Equation 12-1 as

$$n_o = 10.51 \frac{[(0.4)(0.6) + (0.6)(0.4)]}{(0.6 - 0.4)^2} = 126.1$$

Rounding 126.1 to the next highest integer, f is equal to 127, and n is computed as 126.1 x 130/128 or 128.1. Therefore 129 samples in each random sample, or 258 total samples, are needed to detect a difference in proportions of 0.2. Since these are proportions, the result means that the total count in the sample must be at least 129. For example, to detect the above difference in the proportion of dominant taxon (e.g., benthic macroinvertebrates or fish) of two sites, at least 129 individuals must be counted and identified in each estuary.

The example illustrates that a statistically significant difference can be easily detected in proportions if sufficient individuals are sampled. However, it is doubtful that a difference between 40% and 60% in dominant taxon is biologically meaningful.

Example 2—Sample size calculation for comparing population mean abundance

To detect a difference of 20 in mean abundance with a two-sided test. The standard deviation, s, was estimated as 30 for both samples based on previous studies; α was selected as 0.05; and 1-β was selected as 0.90. Substituting these values into Equation 12-2 yields

$$n_o = 10.51 \frac{(30^2 + 30^2)}{20^2} = 47.3$$

Rounding 47.3 to the next highest integer, f is equal to 48, and n is computed as 47.3 x 51/49 or 49.2. Therefore 50 samples in each random sample, or 100 total samples, are needed to detect a difference of 20.

operational QC. Replication of samples at a randomly selected subset of field sites (usually, 10 percent of the total number is considered appropriate) is used to estimate precision, and representativeness of the samples and the methods. Splitting samples into subsamples can be used to check precision of the methodology, and reprocessing of finished samples is used to check accuracy of laboratory operations.

12.3.1 Field Operations

For the field operations aspect of an ecological study, the major QC elements are: instrument calibration and maintenance, crew training and evaluation, field equipment, sample handling, and additional effort checks. The potential errors in field operations range from personnel deficiencies to equipment problems. Field notes are integral to the documentation of

Figure 12-3

Six qualitative and quantitative data characteristics usually employed to describe data quality.

> ▸ **Precision** - The level of agreement among repeated measurements of the same characteristic.

> ▸ **Accuracy** - The level of agreement between the true and the measured value, where the divergence between the two is referred to as bias.

> ▸ **Representativeness** - The degree to which the collected data accurately reflect the true system or population.

> ▸ **Completeness** - The amount of data collected compared to the amount expected under ideal conditions.

> ▸ **Comparability** - The degree to which data from one source can be compared to other, similar sources.

> ▸ **Measurability** - The degree to which measured data exceed the detection limits of the analytical methodologies employed; often a function of the sensitivity of instrumentation.

activities and can be a potential error source if incorrect recording occurs. Training is one of the most important QC elements for field operations. Establishment and maintenance of a voucher specimen collection should be considered for biological data. Transcription errors during data entry can be reduced with double data entry. Table 12-3 gives examples of QC elements for field and laboratory activities.

12.3.2 Laboratory Operations

The QC elements in laboratory operations include sorting and verification, taxonomy, duplicate processing, archival procedures, training, and data handling. Potential error sources associated with sample processing are best controlled by staff training. Controlling taxonomic error requires well-trained staff with expertise to verify identifications. Counting error and sorting efficiency are usually the most prominent error considerations; they can be controlled by training and by duplicate processing, sorting, and verification procedures.

12.3.3 Data Analysis

Errors can occur if inappropriate statistics are used to analyze the data. Undetected errors in the data base or programming can be disastrous to interpretation. Problems in managing the data base can occur if steps are not taken to oversee the data handling, analysis, and summarization. The use of standardized computer software for data base management and data analysis can minimize errors associated with tabulation and statistical analysis. A final consideration is the possible misinterpretation of the findings. These potential errors are best controlled by qualified staff and adequate training.

12.3.4 Reporting

QC in reporting includes training, peer review, and the use of a technical editor and standard formats. The use of obscure language can often mislead the reader. Peer review and review by a technical editor are essential to the development of a sound scientific document.

Table 12-3 Example QC elements for field and laboratory activities

Project Activity	QC Element	Evaluation Mechanism
Field Sampling	Replicated samples at 10% of sites by same field crew.	Calculate relative percent difference (RPD) of index value or individual metric score
	Replicated samples at one to two of total sites by different field crew using same methods.	Calculate RPDs as above; use to evaluate consistency and bias.
Physical Habitat Assessment (Qualitative)	Ensure appropriate training and experience of operators; multiple observers.	Resume or other documentation of experience; discuss and resolve differences in interpretation.
Physical Habitat Assessment (Quantitative)	Replicated measurements at 10% of sites.	Calculate RPDs between replicate measurements; compare to preestablished precision objectives.
Laboratory: Sample Sorting	Sample residue checked for missed specimens to estimate sorting efficiency; check completed by separate lab staff.	Calculate percent recovery; compare to preestablished goals.
Laboratory: Sample Tracking	Logbook with record of all sample information.	Not applicable.
Laboratory: Taxonomic Identification	Independent identification and/or verification by specialist; ensure appropriate and current taxonomic literature available; adequate training and experience in invertebrate identifications; reference collection; exchange selected samples/specimens between taxonomists.	Calculate percent error; compare to preestablished goals.
Data Management	Proofreading; accuracy of transcription.	All transcribed data entries compared by hand to previous form—handwritten raw data, previously computer-generated tables, or data reports.
Data Analysis	Hand-check of reduced data.	For computer-assisted data reduction, approximately 10% of reduced data recalculated by hand from raw data to ensure integrity of computer algorithm.
	Appropriate statistics; training.	Review by statistician or personnel with statistical training.

Case Study: Optimization of Benthic Sampling Protocols: gear, mesh size, replicates

Ferraro et al. (1994) studied the cost-effectiveness of several alternative marine benthic sampling protocols, including sampling gear, mesh size (0.5-mm or 1.0-mm), and number of replicates (1-10), in southern California waters. Alternative sampling gear was:

- 0.1-m^2 van Veen grab
- 0.06-m^2 van Veen grab
- 0.1-m^2 van Veen grab subsampled by 1-6 core samples, 50-300-cm^2 total area subsampled.

Laboratory processing time was recorded for each sampling alternative. Twelve measures of community structure were examined. Results showed that the power of detecting differences between sites did not increase greatly for more than 4 replicates. Optimum cost-effectiveness was achieved with 5 core subsamples (250-cm^2) of 0.1-m^2 grabs, replicated 4 times at each site (Ferraro et al. 1994).

Case Study: Optimization of Benthic Sampling: Seasonal sampling, trend detection

Alden et al. (1997) examined seasonal and annual trends in estuarine benthic macroinvertebrates community measures (diversity, total abundance, biomass, % opportunities). Samples were taken seasonally (4 x per year) from 16 Chesapeake Bay sites for 9 years. Long-term trends were examined by season, and the power of detecting trends was examined for alternative sampling frequencies of 1 season, 2 seasons, or 4 seasons per year. Finally, reference and impaired sites were compared among seasons to determine if some seasons yield greater power of detection of impairment than other seasons.

Trends in indicator values were apparent and detectable in all seasons. Although 4-season sampling yielded the greatest power of trend detection, it was only marginally better than 2-season sampling and 1 season sampling. In general, summer sampling was most sensitive and yielded the greatest power, allowing detection of trends of 4%-7% change per year in abundance, diversity, and % opportunist metrics over the 9 year period. Biomass was much more variable: the minimum detectable trend was approximately 20% change per year for summer-only sampling (Alden et al. 1997).

Chapter 13
Case Studies

In conjunction with the drafting of this guidance manual, the USEPA Biocriteria Program has also supported or assisted in the support of projects around the country to evaluate estuarine and coastal marine survey methods and to develop metrics appropriate for use in different settings. These studies were conducted prior to the creation of this guidance document. Each case study exemplifies a section within the guidance. This chapter summarizes studies conducted in the Pacific Northwest, Gulf of Mexico, and along the Middle and South Atlantic coasts.

Some of the material presented here also appears in the body of this text and the information which follows expands on that discussion. Further, the principal investigators or other contact in each instance are listed with their addresses and phone numbers should the reader desire to comment or request more information.

13.1 Puget Sound - Development of Trawl-Based Tools For the Assessment of Demersal Fauna (Macroinvertebrates and Fishes): A Puget Sound Pilot Study

The relationship between pollution and the health and status of marine benthos are being studied in the Puget Sound region of Washington state (Figure 13-1). Detailed sediment data, including information on chemistry, toxicity and infaunal populations, are being collected for the Puget Sound Ambient

Monitoring Program (PSAMP) as well as for various urban bay, dredge disposal, and Superfund action programs. The PSAMP has quantitatively defined population patterns of demersal fishes, but has not defined those patterns for other demersal fauna such as macroinvertebrates. Nor has any program developed data to explain how the fauna are responding to the environmental stresses associated with a contaminated substrate.

13.1.1 Study Objectives

The Puget Sound study (Eaton and Dinnel 1993, Eaton 1994, 1995) was initiated in 1993 to document demersal populations in their entirety, and to attempt to relate the resulting biological information to sediment chemistry, toxicity, and infauna. The pilot study, funded by USEPA, began by assessing the utility of using two different trawls to quantitatively define demersal populations at a given point in time. Using the resulting documented population patterns and comparisons between reference and contaminated areas, the study objectives were ultimately to:

▸ Gain a greater understanding of how demersal populations are being affected by pollution and habitat degradation;

▸ Determine which patterns reflect environmental stress;

▸ Develop metrics (biological measures) which would help to build a biological index for the rapid

---

Figure 13-1

General location of the case studies.
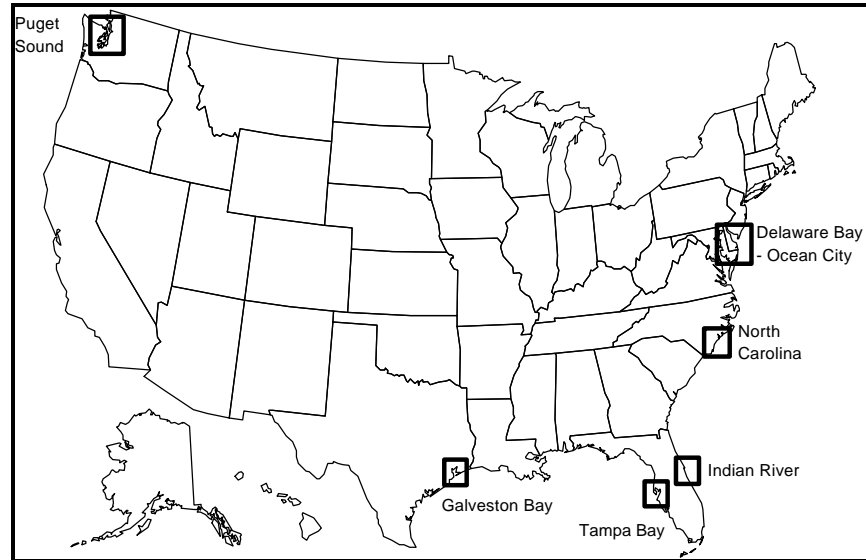
and economical assessment of stress in subtidal biotic marine communities.

13.1.2 Study Methods

Beam and otter trawls were used to sample Puget Sound demersal fauna in 1993 and 1994. A 3-m beam trawl (with a tickler chain attached in front of the net) towed at 1.5 knots proved to be very effective for sampling most demersal invertebrates and small or juvenile fishes. A 7.6-m Southern California Coastal Water Research Project otter trawl, towed at 2.5 knots, was best suited for sampling larger and more mobile marine fishes and invertebrates. All trawl catches were held in tubs of running seawater following capture, and fauna were subsequently sorted, identified, counted, measured and weighed. All organisms were released on station.

In the first year of the pilot study (1993) sampling focused on two of the Tacoma Waterways, the contaminated Hylebos Waterway (a Superfund site) and the adjacent, less-impacted Blair Waterway. Sampling in 1994 compared another Tacoma Waterway and Superfund site, Thea Foss (City) Waterway, with a cleaner and more natural reference

condition (six miles to the north) in Quartermaster Harbor on Vashon Island. Bottom depth, sediment grain-size analysis (either historical information or wet-sieving technique for percent fines), bottom temperature, and salinity data were recorded for all stations to insure meaningful pairing of sites.

Spatial coverage of the study area was determined using a stratified random design. The Hylebos and Blair Waterways were divided into four strata, the Thea Foss into three strata, and Quartermaster Harbor into two. One station was located in each stratum, except in mid-Quartermaster Harbor with two stations one being an historical sediment monitoring station. A second station was placed in the mid-Quartermaster Harbor stratum to compare the variability in results between two stations of close proximity with similar depth and sediment grain-size.

Evaluation of ten consecutive and seven non-consecutive otter trawl replications in 1993 led to the conclusion that four or five otter trawl replications were needed to quantitatively define the demersal fish community at a given station, and that the replications should not recur in

less than four hours. The 1994 sampling design incorporated these recommendations, utilizing a sampling effort of five otter trawl and three beam trawl replications per station.

The extensive data set resulting from the trawl surveys was entered into computer spreadsheets as catch files, and was sorted and statistically analyzed for patterns and relationships. The null hypothesis for the pilot study was that the contaminated and non-contaminated sites were not significantly different for the parameters measured (i.e., fish abundance, biomass, mean individual weights, diversity and evenness). Data comparisons were tested for statistical significance using either a parametric test (i.e., Student's two-tailed, two-sample t-test either paired or independent) or a non-parametric test (i.e., two-sample Kolmogorov-Smirnov Test) depending on the outcome of the test for normality (i.e., one-sample Kolmogorov-Smirnov normality test). Species diversity was calculated using the Shannon-Wiener Index (H') with the natural logarithm, although simple species richness measurements proved to be more statistically significant. Species evenness was measured with Pielou's evenness index (J) and number of species ≥90% of total abundance. Dominance was measured using the dominance ratio, N_{max} / N, where N_{max} = number of individuals of the most abundant species, and N is the total catch.

13.1.3 Study Results

Pilot study results focused on a comparison of the reference and contaminated stations which showed the best match of environmental parameters. Reference station QMH1 in upper Quartermaster Harbor and contaminated station TF1 in the upper end of the Thea Foss Waterway proved to be very similar in depth, sediment grain size, bottom temperature and salinity. A comparison of catch data for the two stations indicated that fish abundance in the reference area was actually lower compared to that found at the contaminated Superfund site (Figure 13-2a), whereas fish biomass was significantly greater at the reference site (Figure 13-2b). This finding indicated that the individual fishes at the reference site must be considerably larger than those found at the contaminated site, and/or that sensitive fish species found at the reference site but not at the contaminated site tended to be much larger than the other fish. Both factors contributed to the differences. Eight of the thirteen fish and invertebrate species common to both sites showed significantly greater mean individual weights at the reference stations, and of the remaining five, only one species was consistently larger at the contaminated sites (Figure 13-2c). Also the cartilaginous fishes (i.e. spiny dogfish, spotted ratfish, and the skates), tentatively classified as sensitive species, were only rarely encountered in the contaminated waterways and were very large compared to the bony fishes.

A preliminary list of tolerant and sensitive fish and invertebrate species was generated for the Tacoma waterways and Quartermaster Harbor fauna based on the pilot study results (Table 13-1). Tolerant species were defined as those whose relative abundance at contaminated sites is significantly greater than or indistinguishable from those species found at comparable reference sites (i.e., site of comparable depth, salinity, dissolved oxygen, sediment grain-size, slope, and density of structures such as eelgrass). Sensitive (intolerant) species, on the other hand, were defined as those species whose relative abundance is

Figure 13-2a

Bony fish abundance and total fish abundance for reference and contaminated sites.

Bony Fish Abundance and Total Fish Abundance for Reference (QMH) and Contaminated (TF) Sites

OTTER TRAWL

P=0.0003

P=0.0003

□ Bony Fish (Osteichthyes) Abundance

■ Bony plus Cartilaginous Fish Abundance

Abundance per 100 m2

QMH 1.1 QMH 1.2 QMH 1.3 QMH 1.4 QMH 1.5   TF 1.1 TF 1.2 TF 1.3 TF 1.4 TF 1.5

REFERENCE   CONTAMINATED

Station and Replication
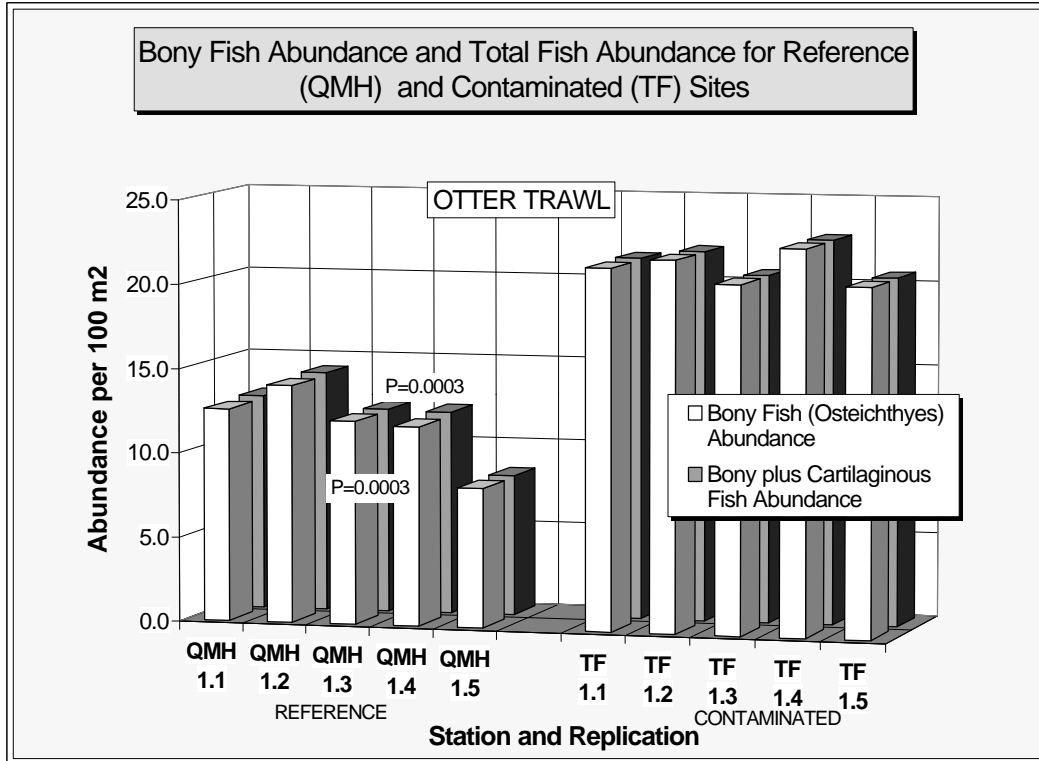
Figure 13-2b

Bony fish biomass and total fish biomass for reference and contaminated sites.

Bony Fish Biomass and Total Fish Biomass for Reference (QMH) and Contaminated (TF) Sites

OTTER TRAWL

P=0.0001

P=0.0001

□ Bony Fish (Osteichtheys) Biomass

■ Bony plus Cartilaginous Fish Biomass

Biomass per 100 m2

QMH 1.1 QMH 1.2 QMH 1.3 QMH 1.4 QMH 1.5   TF 1.1 TF 1.2 TF 1.3 TF 1.4 TF 1.5

REFERENCE   CONTAMINATED

Station and Replication
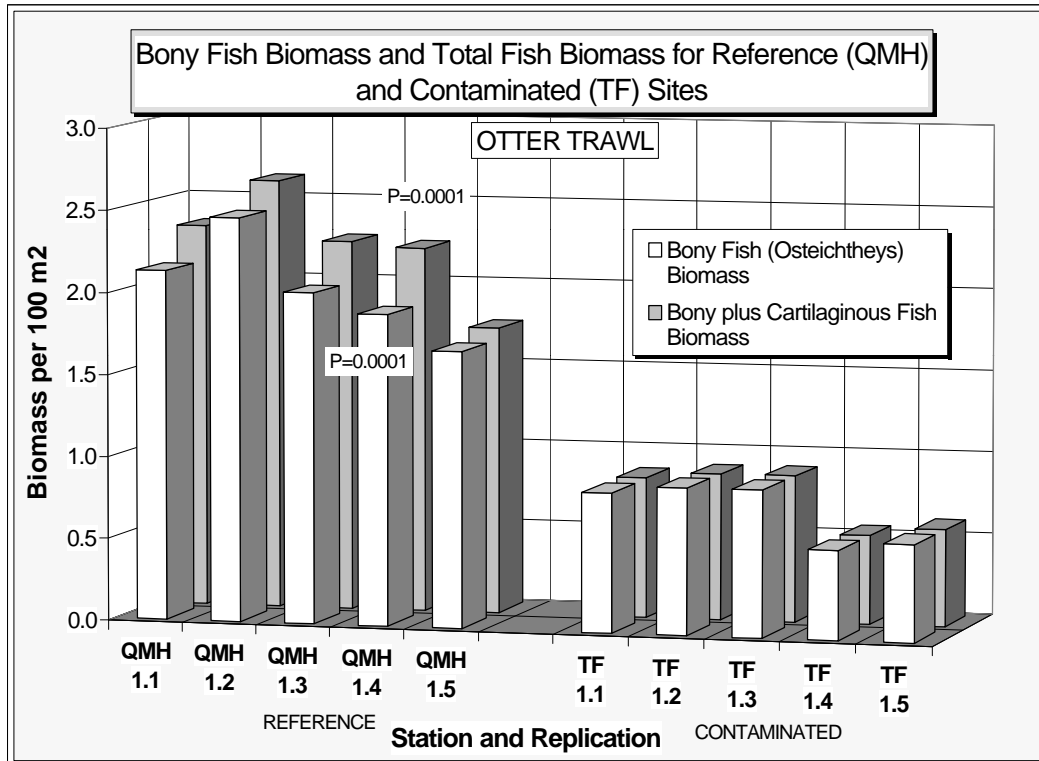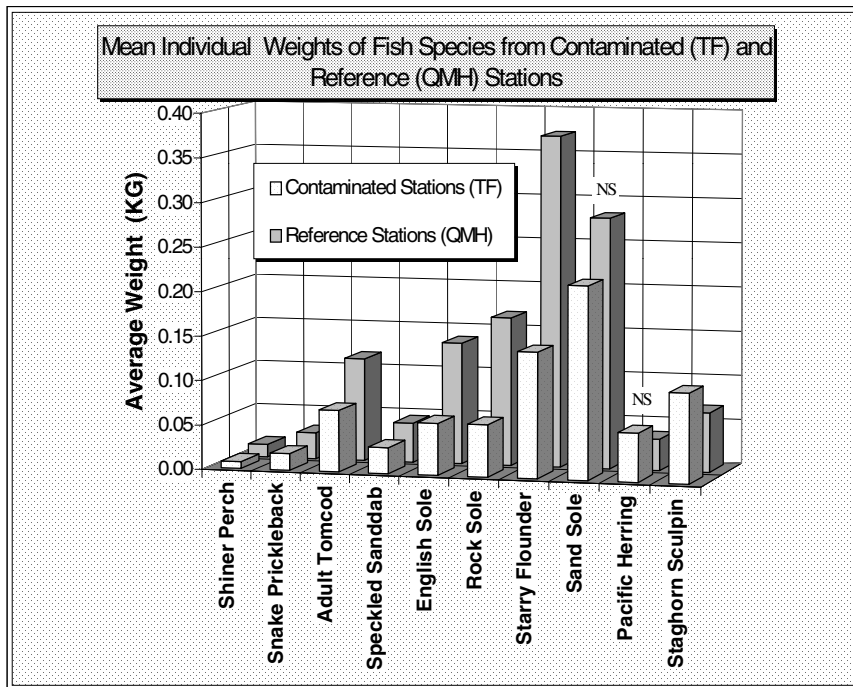
Figure 13-2c

Mean individual weights of fish species from contaminated and reference stations.

significantly greater from a reference area than from a comparable contaminated site.

The sensitive species index, derived from the proportion of sensitive species abundance or biomass to the total of sensitive plus tolerant species, was applied to the pilot study catch data. Index results showed significant differences for all comparisons (i.e., fish abundance and biomass, and fish plus invertebrate abundance and biomass between contaminated and reference sites). The results suggested that such an index, if tested independently for annual and seasonal variation, could be very useful in tracking recovery of an area after cleanup or remediation, or to help classify impacted sites relative to the benchmarks established through the biocriteria.

Pilot study results also indicated that fish species richness and fish species evenness were useful measurements in the site discrimination process.

Although no difference was found in species richness using the beam trawl sampling method, otter trawl catches indicated that fish species richness was notably greater at the reference site (16 species) than at the contaminated station (11 species). When statistically examined on a trawl-by-trawl basis (i.e., using the mean number of fish species per sample), fish species richness was significantly greater at the reference stations. Fish species evenness, as measured by the number of fish species $\geq 90\%$ of total abundance, was also significantly higher at the reference stations, both when paired with the Thea Foss stations, and when compared as a whole. External abnormalities or anomalies, such as fin erosion or skin tumors, were extremely rare at all stations during both study years, thereby suggesting that it may not be a useful indicator of environmental stress.

The results of the first year of sampling indicated that raw or averaged abundance data were not useful in differentiating contaminated and reference sites. This discovery led to an increased effort in recording biomass data during the second study year, and to the inclusion of a more natural reference condition. Results of the second year of sampling emphasized the

Table 13-1. A preliminary list of tolerant and sensitive fish and invertebrate species from the Tacoma Waterways and Quartermaster Harbor.

FISH	
Tolerant Species	**Sensitive Species**
English sole Sand sole Flathead sole Pacific tomcod Shiner surfperch Snake prickleback Pacific staghorn sculpin	Spiny dogfish Spotted ratfish Longnose skate Rock sole Starry flounder Speckled sanddab Pile surfperch Striped surfperch Bay goby Blackbelly eelpout Bay pipefish Plainfin midshipman
INVERTEBRATE	
Pandalus danae: coonstripe shrimp *Crangon spp.*: sand shrimp *Cancer gracilis*: purple cancer crab *Cancer productus:* red rock crab *Cancer magister*: Dungeness crab *Lophopanopeus bellus*: crab *Evasterias troschelli*: mottled seastar *Metridium senile*: plume anemone	*Cucumaria miniata*: sea cucumber *Cucumaria piperata*: spotted sea cucumber *Pentamera populifera*: crescent sea cucumber *Parastichopus californica*: edible sea cucumber *Solaster stimpsoni*: sunstar *Pagurus spp.*: hermit crabs *Nassarius mendicus*: snail

ecologically important fact that the reference sites, despite fewer or equal numbers of fishes, supported more than twice the fish biomass than the contaminated site. Almost every fish species common to both areas was significantly larger, and fish species richness and evenness were significantly higher at the reference site.

The sensitive species index proved to be useful in differentiating sites. The identification of sensitive (intolerant) and tolerant demersal marine species is in its infancy, due in part to the paucity of data on demersal marine communities and the lack of quantitative sampling methods. With the development of these sampling techniques, the pilot study demonstrated that information on the demersal fauna should be included in

any future ecologically-based indexes of pollution. Candidate attributes of demersal fauna which warrant further study are listed in Table 13-2.

Primary Contact: Charles Eaton
Bio-Marine Sciences
2717 3rd Ave. N
Seattle, WA 98109
206-282-4945

Table 13-2. Candidate attributes of demersal fauna showing significant differences in the present study.

Candidate Metrics	Preliminary Expectation from Present Study: **IMPAIRED SITE**	Range of Values (and mean) from Present Pilot Study: **IMPAIRED SITE:** Thea Foss Wty.	Range of Values (and mean) from Present Pilot Study: **REFERENCE SITE:** QM Hbr.
Total **Fish Abundance** per 100 m², 4-6 m. Depth. (10-15 m.—no difference).	Elevated or no difference	TF 1: 20.6 to 22.7 mean = 21.4	QMH 1: 8.3 to 14.2 mean = 11.9
Total **Fish Biomass** per 100 m²	Reduced	0.54 to 0.97 kg (mean = 0.73)	0.81 to 4.82 kg (mean = 2.64)
Fish Species Richness (Number of fish species)	Reduced	per tow: 6-13 cumulative:10-14	per tow: 11-18 cumulative:15-18
Fish Species Evenness (Number of Fish Species $\geq$ 90% of Total Abundance)	Reduced	3-6 (mean = 3.7)	4-7 (mean = 5.2)
Mean Individual Weight and Size of all Species (except Pacific Herring and Staghorn Sculpin)	Reduced	e.g. English Sole 45 to 114 g (mean = 75 g.)	e.g. English Sole 125 to 484 g. (mean=204 g.)
Tolerant Species Abundance and Biomass (English Sole, Pacific Staghorn Sculpin, Pacific Tomcod, Shiner Surfperch, Snake Prickleback, Purple Crab, Mottled Seastar, Plum Anemone)	Elevated or (in some cases) no difference	e.g. English Sole 2.8-10.1 (5.3) per 100 m² e.g. juv. Tomcod 0.39 to 9.87 (7.3) e.g. Purple Crab 4.7 to 56.0 (22.8)	e.g. English Sole 0.2 to 6.1 (2.3) per 100 m² e.g. juv. Tomcod 0 to 0.69 (0.4) e.g. Purple Crab 0.0 to 6.3 (2.7)
Sensitive Species Abundance and Biomass (Bay Goby, Starry Flounder, Rock Sole, Cartilaginous Fish, Sea Cucumbers)	Reduced	e.g. Bay Goby 0.4 to 3.9 (1.7) per 100 m² e.g. *Cucumaria piperata* Zero	e.g. Bay Goby 28.6-41.9 (36.5) per 100 m² e.g. *Cucumaria piperata* 0.0 to 5.9 (2.5)
Sensitive Species Index (proportion of sensitive to sensitive + tolerant)	Reduced	0.016 to 0.098 mean = 0.049	0.075 to 0.787 mean = 0.357

13.2 Galveston Bay - Development of a Rapid Bioassessment Method and Index of Biotic Integrity For Coastal Environments: Northwestern Gulf of Mexico Pilot Studies

13.2.1 Study Objectives

A study was conducted on selected streams and bayous within Galveston Bay, Texas (Figure 13-1) coastal ecosystems, in order to characterize the expected fish assemblages of various types of waterbodies (with varying water and habitat quality) (Guillen 1995a). A second study objective was to develop a prototype rapid bioassessment technique similar to the Index of Biotic Integrity for the northwestern Gulf of Mexico. In order to meet the second objective, several criteria for the development of the method had to be met. First, the method had to be ecologically relevant, that is any metric or ranking system had to directly relate to ecological function and structure. Secondly, the method had to be taxonomically simple or kept to the broadest taxonomic/functional group of organisms that provide the most information. The methods also had to be simple (in terms of equipment, labor, and analysis) cost effective, easily standardized, subject to easy replication, and adaptable to a variety of environments.

13.2.2 Study Methods

The sampling design consisted of five bayous classified according to the potential for anthropogenic impact; i.e., urban versus rural, impaired versus unimpaired and salinity effects; i.e., lower portions of tributaries versus upper portions. Oyster Bayou, Dickinson Bayou, Texas City Hurricane Canal, Highland Bayou Diversionary Canal, and Cedar Lakes Creek were selected to fulfill these criteria. Oyster Bayou is a minimally impaired coastal bayou located in the middle portion of Galveston Bay and flowing south to East Bay. Oyster Bayou stations were characterized by a silty clay substrate. The moderately impaired Dickinson Bayou is located in the northeastern portion of Galveston County. Dickinson Bayou is characterized by sandy to silty clay substrates and is impaired by both point and nonpoint sources. The Texas City Hurricane Canal is an industrial canal that flows into the Texas City ship channel, and receives industrial and stormwater discharges. The majority of the canal banks possess a steep slope, and little bank vegetation, and the southern shoreline is an artificial levee. The Highland Bayou Diversionary Canal is an artificial waterbody created by the Army Corps of Engineers in 1983. The canal was created by channelization of the upper reach of Highland Bayou proper and construction of an earthen dam directly below the channelized portion, in order to reroute water through a dredged canal into Jones Bay. The canal is tidally influenced and receives effluent discharge from municipal wastewater treatment plants and runoff from surrounding agricultural grazing and pasture lands. Cedar Lake Creek is a minimally impaired rural bayou which extends 24-miles from its origin at the intersection of Cedar Lakes to the Gulf Intracoastal Canal. There are no active discharges in the watershed, however, an oilfield is present at its upper reaches. Predominant land use in the area is cattle grazing and the San Bernard Wildlife Refuge.

To summarize, two minimally impaired bayous (Oyster Bayou and Cedar Lakes

Creek) and three impaired waterbodies (Highland Bayou Diversionary Canal, Texas City Hurricane Canal and Dickinson Bayou) were surveyed during the study period. The impaired sites included two that were influenced by residential and municipal wastewater, and one effected by industrial effluent. Two of the waterbodies were highly channelized and/or man-made. Site investigations involved seasonal quarterly surveys made at all stations within each watershed. Sampling was conducted during summer, fall, and winter 1991; spring and summer 1992; and winter, spring, summer, and fall 1993.

In order to evaluate the relationship between water quality and fish communities, various hydrological, habitat, and biological data were collected concurrently. Qualitative habitat measurements including primary and secondary tributary depth, width, substrate type, and shoreline vegetation were noted at each station. A rapid field method for the evaluation of percent sand in sediments was also used to evaluate effects of sediment size on nekton populations. Measurements of surface and bottom temperature, dissolved oxygen, conductivity, salinity and pH were made. Surface water samples were also collected for the determination of total organic carbon, fecal coliforms, total and orthophosphate, nitrates, total ammonia, total suspended solids, and chlorophyll *a*. Individual water chemistry and habitat values were plotted against seasons and stations to evaluate temporal and spatial patterns.

In addition, Pearson's correlation coefficients and stepwise and direct discriminant analyses were used to determine the relationship between the variables and clustered groupings of stations. The analyses provided another tool for investigators to evaluate the relative influence of physicochemical variables on coastal nekton communities. Survey results showed that the majority of water quality variables were within previously documented tolerance limits of estuarine fish.

Nekton (fish and macrocrustacea) were collected using experimental gillnets, trawls, and seines. Gillnets were 200 x 8-ft experimental monofilament nets with eight panels of varying mesh sizes (0.5-4-in mesh). Seine collections (five replicates of 25-ft hauls) were made using a 15 x 4-ft common minnow seine with 1/8-in square mesh nylon netting. Trawls were made at main channel stations in each watershed, using a 10-ft otter trawl with 1-in mesh in the wings and 1/4-in mesh in the cod end. Four replicate trawls (five minute tows, each) were made at each of the mainstream stations. Nekton collected via all sampling methods were identified to the lowest possible taxon, enumerated, and measured.

13.2.3 Study Results

Several biological metrics were considered during the pilot study based on historical usage and recent recommendations in the literature. Community metrics generated from pilot study data included: total catch, log-transformed total catch, number of nekton taxa, Shannon-Wiener diversity index, Pielou evenness index, total number of taxa making up 90% of the catch, dominance ratio (ratio of most abundant species/total catch), number of crustacean species; number of "bottom taxa"; i.e., sciaenids, flatfish, blue catfish; number of predatory species; number of "minnow" taxa; i.e., Poeciliids and/or cyprinodonts; number of goby taxa; proportion of total catch as bay anchovy; proportion of total catch as

"shad"; i.e., clupeids and engraulids, proportion of total catch as poeciliids; proportion of total catch as Penaeid shrimp; and proportion of total catch as palaemonid shrimp; i.e., grass shrimp. The rationale for each proposed metric is provided in Table 13-3.

Trawl, seine, and gillnet results were utilized in a cluster analysis, and then subjected to stepwise discriminant analyses. Observed seasonal and spatial patterns and/or temperature and salinity related correlations were used to determine whether seasonal or salinity-adjusted metrics were needed. In addition, the decision to include data from "impaired" sites in the derivation of metrics was also evaluated using these analyses. If initial statistical analyses failed to show differences between the reference sites and impaired sites, then all sites were pooled for derivation of metrics.

Due to the strong seasonal and/or spatial trends observed in various metrics using the seine data, cumulative percent distributions of each candidate metric were calculated by season for minimally impaired watersheds. Results of the distributions are presented in Table 13-4. Metrics were adjusted where distributions indicated truncated values using the following approach. If the distribution line could be extended to the 15th or 85th percentile value without crossing the Y axis, then that estimated value was used. If it could not, the metric was not used during that season and/or the metric rating was adjusted to reflect only two conditions (e.g., normal and excellent). This same procedure was used to derive a proposed metric system using trawl data (Table 13-5). A proposed list of prototype metrics using gillnet catch data was developed (Table 13-6); however, since gillnet design and deployment is variable, it may be

difficult to compare metrics derived from the pilot study with other studies. Pilot study results indicated that it seems feasible that a prototype estuarine bioassessment system based on nekton community collections can be used to successfully document impacts from pollution. Analysis of potential metrics through discriminant analysis, graphical evaluation of cumulative distribution, frequencies and correlation analyses yielded the proposed metrics listed in Tables 13-4, 13-5 and 13-6. The categories utilized in the framework of the proposed system were based on the following protocol. Depending on the metric, those values less than the 15th percentile were categorized as "concern". The interquartile values; i.e., 15-85th percentile were categorized as "normal", and values exceeding the 85th percentile were classified as "excellent". In some cases where high metric values denoted degraded conditions, the inverse of the proposed scheme was used; i.e., ‹15th percentile = "excellent". The classification system was based primarily on statistical distributions of the observed data. Where data was insufficient, a "not recommended" disclaimer was listed.

It was difficult to single out one water quality variable as having the most influence on community structure and the proposed metrics. Therefore, a conservative approach was taken by grouping by season and utilizing all data across various salinity levels.

The proposed Index of Biotic Integrity (IBI) metrics derived from the pilot study would be most confidently applied to situations where salinities range from 1-25-ppt. Continued calibration of this system with additional data sets is needed. The proposed metrics need to be evaluated against independent data sets including those in high (>25-ppt) salinity regimes.

Table 13-3. Rationale for the inclusion of proposed nekton community metrics.

Metric	Rationale
Total Catch	Total abundance is a rough measure of the total community population and as such gives no information on individual species population levels. Low abundances can be caused by various stressors. It should be noted that high abundances caused by individual opportunistic species can also indicate a disturbed community.
Log Total Catch	Due to the inherent variability of populations, the patchiness of fish schools and previously observed distributions of fish, many ecologists feel that the log-normal distribution fits the distribution of nekton populations better. Therefore log total catch may be a more appropriate indicator of total population levels. In order to handle zero catches, however, a log (catch + 1) transformation is needed.
Total Number of Nekton Taxa	The species richness of any community is extremely important. Reductions in species number may indicate an overall reduction in available habitat or the presence of environmental stressors. This may be due to the avoidance or death of sensitive species in an area. The number of taxa collected is a relatively economical measure. On a relative scale it is the cheapest information obtainable from catch data.
Cumulative Number of Nekton Taxa	The cumulative number of taxa is somewhat different than the total number of taxa in that it reflects the upper limit of the number of taxa one would expect to collect within a single replicate sample. Large discrepancies between mean and cumulative number of species may indicate high variability in habitat or distribution of species. Like the total number of taxa metric a low cumulative number of taxa can reflect limited habitat and/or the presence of environmental stressors.
Total Number of Fish Taxa	This metric is closely related to total nekton species numbers. However, it was added to address situations where only fish data is tabulated.
Nekton Species Diversity	The Shannon-Wiener diversity function was selected to evaluate nekton communities. This commonly used function (H') was developed to incorporate the two most important components of diversity, namely richness and evenness. Species richness is normally tabulated. However, species richness alone provides no information on how evenly individuals are distributed among species. The majority of communities studied by ecologists show a log-normal pattern of species abundance in which a relatively few species possess a rather large number of individuals and a rather large number of species possess few numbers of individuals. A diverse community is one in which species number and evenness are maximized. One problem with the use of H' is the fact that various combinations of species numbers and evenness can yield the same answer. Therefore diversity indexes should only be evaluated in the presence of species richness and evenness.

METRIC	RATIONALE
Nekton Evenness	One factor that influences diversity directly is the evenness of the distribution of organisms between species. Populations possessing high numbers of taxa but with highly uneven distributions between taxa (e.g. highly dominant taxa) may reflect underlying habitat limitations, stressors or seasonal patterns. One of the most popular indexes used by marine biologists is the Pielou's evenness index (J). This index is defined as: $J = H'/\ln(S)$, where H' is the Shannon-Weiner index, ln is the log base (e) and S is equal to number of taxa This index expresses H' relative to the maximum value that H' can obtain when all of the species in the sample are perfectly even with one individual per species.
Number of Nekton Taxa = 90% Catch	This index is the number of taxa that together add up to at least or exceed 90% of the total catch. This is another measure of evenness. High values would indicate a community in which there is no clear dominant taxa. This index is influenced by the same factors which effect the evenness index.
Nekton Dominance Ratio	This has also been referred to as the Berger-Parker index. This is the ratio N_{max}/N, where N_{max} = number of individuals present in the most abundant taxa, and N is the total catch. This equation is computationally simple and can be easily programmed into spreadsheets. In addition, it is intuitively easy to understand. High dominance reactions reflects dominance of the community by a few individuals which relates to an uneven distribution of individuals within taxa resulting in poor diversity. This may be related to potential stressors and other factors cited under the discussion of Pielou's evenness.
Number of Crustacean Nekton Taxa	The number of crustacean taxa present in the nekton is largely a function of 4 principle groups. The first group are crab species including blue crab, *Callinectes sapidus*. The second group includes seasonally dominant groups of Penaeid shrimp which migrate into tidal creeks and bayous as postlarva and juveniles. The third group includes resident species of grass shrimp, genus *Palaemonetes*. The final group include freshwater prawns, genus *Macrobrachium*, and crayfish genus *Procambarus*. The presence of crustacean taxa indicates a healthy population of benthic herbivores and omnivores which serve as the primary food source for many estuarine fish. In addition, crustaceans are especially sensitive to organic pesticides.
Number of Predatory Fish Taxa	Predatory fish were defined as fish in the family Carangidae, Scombridae, and the genera *Paralichthys*, *Lepisosteus*, *Micropterus*, *Cynoscion*, *Morone* and the species *Sciaenops ocellatus*, *Synodus foetens* and *Elops saurus*. These species represent individuals at the top of the food chain. Impacts to other species they depend on may reduce these predators indirectly. In addition, through the process of biomagnification predators are more likely to bioconcentrate high levels of pollutants found in the lower portions of the food chain.

Table 13-3 (Cont'd). Rationale for the inclusion of proposed nekton community metrics.

METRIC	RATIONALE
Number of "Minnow" Taxa	The number of "minnow" taxa include resident species of Cyprinodontidae and Poeciliidae. These two groups of small fish represent the majority of resident species inhabiting marsh and shallow water environments. The majority of these species are not normally found offshore or in deeper waters due to predation. A high number of these taxa may reflect habitat suitability of a particular location to resistant species. Since these species are largely non-migratory, their presence or lack of may indicate long-term environmental perturbation. In contrast, high populations of these species may correlate the absence of larger predators and/or the presence of marginal habitat unsuitable for other less tolerant taxa.
Number of Goby Taxa	This is another group of resident taxa that are primarily carnivorous, feeding on small invertebrates. In addition, gobies are extremely territorial and tend to stay within a defined area. Most gobies are benthic. Reduced numbers of gobies would indicate localized impacts to habitat, water quality and secondary impacts on food sources namely, epibenthic invertebrates.
Proportion of Nekton Catch as Poeciliids	Poeciliids are a group of fish that are generally extremely tolerant to poor water quality. Notable examples include the mosquitofish (*Gambusia affinis*) and molly (*Poecilia latipinna*). These two species are often found in harsh habitats where few other species live. In addition, they are typically found in areas (e.g., shallow flats) which are difficult for predators to exploit. A predominance of Poeciliids in shoreline communities can therefore indicate degraded conditions and/or lack of predators.
Proportion of Nekton Catch as Penaeid Shrimp	One of the most important species ecologically and commercially are the Penaeid shrimp, including the estuarine white shrimp *Penaeus setiferus*, the brown shrimp *P. aztecus*, and the less abundant pink shrimp *P. duorarum*. Typically these species enter the estuaries as postlarvae. With continued migration they reach tidal creeks and bayous as juveniles and spend the early part of the first year in these areas prior to migrating back to the ocean to spawn. The proportion of catch as Penaeid shrimp may be an excellent ecological indicator. Penaeid shrimp can serve as a metric that addresses the nursery functions of a waterbody. It is also a lower food chain omnivore. Many species are dependent upon this invertebrate for food. Therefore, reduced numbers of shrimp can detrimentally effect the entire nekton community. In addition, since it is an arthropod it may serve as an excellent indicator of secondary impacts associated with pesticide use.

Table 13-4. Proposed seine metrics for use in an estuarine IBI along Texas coast.

Metric	Summer Value	Fall Value	Winter Value	Spring Value	Score
Category A					
(A) Total Catch	<200 200-450 >450	<50 50-400 >400	NA ≤900 >900	NA ≤700 >700	1 2 3
*(A) Log Catch	<.5 4.5-6.0 ≥6.0	<.9 3.9-5.8 ≥5.8	<4.2 4.2-6.4 >6.4	<1.5 1.5-6.3 >6.3	1 2 3
*Prop. Pen. Shrimp	<.01 0.01-0.3 >0.3	<25 .25-.56 >56	NA NA NA	NA ≤0.04 >0.04	1 2 3
Category B					
(B) Prop. Shad	>0.83 NA ≤0.83	>.60 .04-.60 ≤0.04	>.59 NA ≤.59	>.78 NA ≤.78	1 2 3
*(B) Prop. B. Anchovy If Bay A. = 0 then use 'Shad' metric	>0.8 NA ≤0.8	>.52 .04-.52 <.04	>.13 NA ≤.13	>.34 NA ≤.34	1 2 3
*Dominance Ratio	>0.88 0.44-0.88 <0.44	>.65 .40-.65 <.40	>.82 .26-.82 ≤.26	>.78 .27-.78 <.27	1 2 3
Category C					
*(C) Mean #Taxa	<6 6-11 >11	<6 6-10 >10	<6 6-10 >10	<5 5-10 >10	1 2 3
(C) Cum. #Taxa	<10 10-19 >19	<6 6-11 >11	<11 11-18 >18	<11 11-19 >19	1 2 3
(C) Mean #Fish Taxa	<3 3-7 >7	<3 3-7 >7	<3.5 3.5-7 >7	<4 4-8 >8	1 2 3
Total IBI Score					
Concern	5-7	5-7	4-5	7-9	
Normal	8-12	8-12	6-10	10-12	
Excellent	13-15	13-15	11-12	13-15	
Total IBI Score (WHEN INVERTS NOT USED)					
Concern	4-5	4-5	4-5	5-6	
Normal	6-10	6-10	6-10	7-10	
Excellent	11-12	11-12	11-12	11-12	

* Recommended metric; if mean log total catch or total catch = 0, then score = high concern.

Table 13-5. Proposed trawl metrics for use in an estuarine IBI along Texas coast.

Metric	Summer Value	Fall Value	Winter Value	Spring Value	Score
*Prop. Total Catch as P. Shrimp	* ≤0.45 >0.45	<0.42 .42- .83 >0.83	* * *	* ≤0.08 >0.08	1 2 3
*Prop Total Catch as Shad	>0.06 * ≤0.06	>0.08 * ≤0.08	* * *	>0.03 NA ≤0.03	1 2 3
Category A					
*Mean # Nekton Taxa	<1.8 1.8-9.3 >9.3	<4.3 4.3-9.9 >9.9	<4.4 4.4-8.8 >8.8	<4.1 4.1-7.7 >7.7	1 2 3
Mean # Fish Taxa	<2.2 2.2-6.3 >6.3	<4.2 4.2-6.6 >6.6	<2.1 2.1-6.8 >6.8	<1.6 1.6-4.9 >4.9	1 2 3
Total IBI Score					
Concern	4	3	1	4	
Normal	5-8	4-8	2	5-8	
Excellent	9	9	3	9	

NOTE: To avoid problems caused by division by zero use the following formulas: For shrimp and shad proportions let metric value = taxa group catch/(total catch + 1). Alternately if any one replicate total catch = 0, then an IBI score of 'concern' can be given.

NOTE: Avoidance of winter sampling is recommended due to lack of suitable metrics.

* Recommended metric; if mean log total catch or total catch = 0, then score = high concern.

Table 13-6. Proposed gillnet metrics for use in estuarine IBI along Texas coast.

Metric	Assigned Metric Score			
	0	1	2	3
1) Total Nekton Catch	0	1<26	26<47	≥47
2) Category 2 (Pick 1 of the 2 Metrics Listed)				
2a) Number of Nekton Taxa	0	1<7	7<8	≥8
2b) H'	*	<1.25	1.25<1.85	≥1.85
3) Category 3 (Pick 1 of the 2 metrics listed)				
3a) Evenness J	*	<.58	.58<.81	≥.81
3b) Number of Taxa = 90% Total Nekton Catch	*	<4	4<6	≥6
No. Crust. Spp.	*	Not Recom.	Not Recom.	Not Recom.
4) No. Pred. Taxa	*	0	1	>1
No. "Minnow" Taxa	*	Not Recom.	Not Recom.	Not Recom.
5) No. Scaenids/B. Cat. Taxa	*	0	1-2	>2
Total IBI Score (SUM OF ALL 5 METRIC CATEGORIES)	0	5-6	7-13	14-15
Total IBI Rank	HIGH CONCERN	CONCERN	NORMAL	EXCELLENT

Due to the lack of strong correlation between the seine and trawl-derived metrics, it is advisable that future studies use both gear types. Since gillnet derived metrics were least sensitive to water quality fluctuation, and gillnet use is labor intensive and difficult to replicate, gillnetting is the least favored approach for evaluating nekton community health.

A fish health index (FHI) was tested during the pilot study to evaluate its utility in assessing impacts on estuarine fish communities. The FHI methods mirrored those used by the Tennessee Valley Authority (Dycus and Meinert 1993, Dycus 1995). Further evaluation is needed to determine the discriminatory power; i.e., impaired versus unimpaired sites

of the index. Proposed FHI values for Gulf Coast bioassessments are listed in Table 13-7. The FHI proved to be time and cost efficient, and yielded information that was complementary to the IBI.

Table 13-7. Proposed fish health index and condition factors for use in estuarine rapid bioassessments of Texas Gulf coast tidal tributaries.

FISH HEALTH INDEX

Species	Assigned FHI Rank		
	1	2	3
Blue Catfish FHI Score	>70	40-70	<40
Atlantic Croaker FHI Score	>50	30-50	<30
Sea Catfish "Hardhead" (least recommend) score	>30	*	≤30
Spot FHI Score	>61	32-61	<32
AVERAGE SPECIES FHI RANK	1-1.4	1.5-2.5	≥2.6
Total FHI Rank based on average scores of single species	CONCERN	NORMAL	EXCELLENT

CONDITION FACTOR

Species	Assigned Condition Factor (CF) Rank		
	1	2	3
Blue Catfish CF	<.78	.78-.97	>.97
Atlantic Croaker CF	<1.02	1.02-1.15	>1.15
See Catfish "Hardhead" CF	<.82	.82-1.10	>1.10
Spot CF	<1.25	1.25-1.44	>1.44
AVERAGE SPECIES CF RANK	1-1.4	1.5-2.5	≥2.6
Total CF Rank based on average scores of single species.	CONCERN	NORMAL	EXCELLENT

The application of estuarine rapid bioassessment techniques in studies of Gulf of Mexico coastal environments is warranted, based on results of the pilot study. Several of the methods tested in the study (including seine and trawl based IBI and FHI) would aid water quality and fisheries scientists in their evaluation of water and habitat quality impacts resulting from human activity. The pilot bioassessment methods meet the requirements for inclusion in routine monitoring programs including: low cost, low effort, readily obtainable equipment, relatively easy taxonomy, and replication of effort which is suitable for statistical analyses.

Primary Contact: George J. Guillen, U.S. Fish & Wildlife Service
1655 Heindon Rd.
Arcata, CA 95521
707-825-5109
george_guillen@fws.gov

13.3 Tampa Bay - Development of a Community-Based Metric for Marine Benthos: A Tampa Bay Pilot Study

13.3.1 Study Objectives

State biological criteria in Florida have been set at a 25% decrease in Shannon-Wiener diversity of benthic communities in test versus reference sites. Input data have been the sum of three ponar grab samples per site; however, evidence has suggested that these methods and criteria are not sensitive enough. Pilot studies in the Tampa Bay area (Figure 13-1) have tested a process of classifying organisms according to their sensitivity or tolerance to pollution, and developing an index (the Farrell Epifaunal Index) value for test and reference sites (Farrell 1993a). The pilot study used biological data from areas surrounding treatment plant outfalls in the index calculations, in order to detect differences between test and reference sites that were not evident using the state criterion of a 25% decrease in diversity.

13.3.2 Study Methods

Water quality and benthic data were developed from a 1992 short-term study of the effects of three small package plants on the seagrass communities at Fort Desoto Park in Tampa Bay, Florida. Three control stations were located on Joe Island on the southern shore of Tampa Bay, and an additional station was located on a small island adjacent to Fort Desoto (that was presumably under the potential influence of the farfield effects). Two sampling sites were located at each station, one on the shoreline (end of pipe) and a second 50-m offshore. Four petite ponar replicates were collected at each site; however, only three were analyzed for

macroinvertebrates. This process was consistent with Florida's biological integrity standard as defined in the Florida Administrative Code. After the ponar samples were collected, macroinvertebrates were also sampled at each location using a modified Renfro Beam Trawl towed for a distance of 4-m.

The Renfro Beam Trawl is a conical net, open at the large end, which is normally towed over the surface of the substrate. The net is maintained in an open position by attaching it to a rigid pole or beam. The body of the net is constructed of nylon bolting cloth (50 openings/cm^2, which tapers to a plankton net fitted with a removable bucket. The effective swath width of the custom trawl used for the pilot study was 1.25-m. By towing the net over a uniform measured distance, the results were comparative (semiqualitative) and relative abundances of the various species were maintained. The standardized tow length of 4-m effectively sampled approximately 5-m^2 of bottom. Some advantages and disadvantages of using the epibenthic beam trawl are listed in Table 13-8.

In advocating the use of the beam trawl, which predominantly samples the epifaunal and facultative infaunal communities, one basic assumption was made. Provided that the recruitment potential for the individual community components existed, it was assumed that within a given set of natural environmental parameters an expected community of organisms would inhabit any predetermined environmental segment. In estuaries and many other marine environments, populations of different species vary significantly over the seasons and from year to year; however, these variations follow predictable patterns. In Florida, numerical dominance may vary among annual cycles; however, species

Table 13-8. Advantages and disadvantages to using the epibenthic Renfro beam trawl for the sampling of benthos.

ADVANTAGES	DISADVANTAGES
• The epibenthic assemblage is very sensitive to anthropogenic stressors, and this method can be used in both a nearfield and farfield context with equal facility. • Since this method is limited to level bottoms, the total number of common species will be limited (thereby greatly simplifying training). [NOTE: Time required for analysis of three ponar samples was approximately 20-hours, whereas the time required to analyze a pilot study trawl sample was a little less than 10-hours]. • This method lends itself to subsampling which will reduce processing hours and increase cost-effectiveness. • Once initial training is completed, field efforts can be relatively rapid and analytical time can be reduced. • Samples can be sorted qualitatively, and a nonparametric analysis can be applied to provide a method of quick screening.	• The method is restricted to level bottoms. Hard substrates cobble, and emergent vegetation tend to invalidate the method. • In areas of abundant seagrasses or macroalgae beds, sample bulk can be a hindrance and some rough field sorting may be required. • The epifauna tend to be seasonally abundant; therefore, this factor would have to be calibrated into the method if multi-seasonal sampling events are utilized.

composition generally remains stable. Benthic macroinvertebrates, in terms of both density and diversity, reach their peak in Florida during the late winter to early spring (or earlier in the southern part of the state). Population minima for most species occur during the summer months. While they are dramatic, these seasonal cycles can be factored into efforts to establish biocriteria. It is important to consider seasonality because the species which are most sensitive to environmental stress are those which tend to reach their population peaks during periods when water quality factors are both stable and optimal.

The epifaunal and facultative infaunal community was targeted for the pilot study since components of the community appear to be both persistent and very sensitive to environmental stress. Within estuaries and adjacent near-shore areas, physiochemical parameters (e.g., temperature, salinity, dissolved oxygen) will vary significantly over an annual cycle. Sessile and relatively immobile organisms (including most of the infaunal components) have evolved either mechanisms which allow them to tolerate these varying conditions, or breeding cycles which allow them to avoid periods of high stress. The more motile members of the community (including the epifauna and facultative

infauna) have the option of avoidance. During periods of stress, these organisms can move to deeper water or to other areas where stressors are mollified, and then return when conditions improve. When an area is being affected by relatively low levels of anthropogenic stress, only the most sensitive members of the benthic community will respond, and these are found among the epifaunal and facultative infaunal components. It is apparent that a method which is truly sensitive to low levels of pollution must target these components of the benthic community, thereby advocating their use in the Tampa Bay pilot study.

13.3.3 Study Results

A bioassessment approach should be able to not only detect low levels of environmental stress, but it should also be able to detect those stress factors at the earliest possible stages. One approach which has been used successfully in freshwater environments has involved the assignment of the specific index values to various community components; i.e., species, and basing community assessment on the mean index value derived from sampling that community (Lenat 1993). The Farrell Epifaunal Index proposed in the Tampa Bay pilot study was specifically developed for the west coast of Florida, but it should be useful in adjacent areas. The index values represented a somewhat subjective evaluation of the relative tolerance or intolerance to environmental stress. The values were taken from an ongoing effort to assign tolerance values to all marine and estuarine macroinvertebrates identified from the coast of Florida. Information sources have included agency monitoring data, published records, gray literature, and anecdotal information. Wherever possible, all potential stressors including

sensitivity to toxic substances was taken into account; however, the dominant factor for most of the species was the relative sensitivity to dissolved oxygen depression. As a result, the Farrell Epifaunal Index was probably most sensitive to organic pollution and eutrophication with associated wide swings in dissolved oxygen.

The tolerance criteria for the index in terms of dissolved oxygen requirements were as follows:

0   Insufficient data to make an evaluation.
1   Very tolerant. The identified species can withstand short periods of anoxia.
2   Tolerant. The identified species can withstand brief excursions to 1.0-1.5 mgL^{-1}.
3   Slightly tolerant - slightly sensitive. The identified species can withstand brief excursions to 2.5-3.0 mgL^{-1}.
4   Sensitive. The identified species can withstand brief periods below 4.0 mgL^{-1}.
5   Very sensitive. The identified species are basically intolerant of concentrations below 5.0 mgL^{-1}; however, some species may tolerate brief excursions below this provided no other stress factors are involved.

When the components from a sample had been identified, the predetermined tolerance values were assigned to the various species, and a sample Farrell Epifaunal Index value was calculated using the formula:

$$\textit{Farrell Epifaunal Index} = \Sigma \frac{n_i I_i}{N}$$

where: n_i = number of individuals of species i;
         I_i = tolerance value for species i;

N = total individuals from all species used in the calculation.

In a strictly qualitative approach, an index value may be calculated using the formula:

$$\textit{Farrell Epifaunal Index (Qualitative)} = \Sigma \frac{I_s}{N_s}$$

where: I_s = index value for component species s;
N_s = number of species used in the calculation.

Pilot study calculations of Farrell Epifaunal Index values required that the appropriate tolerance value (0-5) be assigned to individual taxa in each sample. The values were then added, and the summation was divided by the total number of taxa utilized from the sample. Taxa with a value of zero were omitted from calculations. Pilot study results (Table 13-9) indicated that the index had been successful at detecting differences between test and reference sites. The resulting Farrell Epifaunal Index will not meet all needs, and is not the only metric that could be applied to beam trawl or similar samples; however, pilot study results indicate that at a minimum it should prove to be an effective screening method.

Primary Contact: Steven Kent, FLDEP,
3319 Maguire Blvd.
Orlando, FL 32803
407-894-7555, ext. 2227
kent_sl@orl1.dep.state.fl.us

Table 13-9. Farrell epifaunal index results for the Fort Desoto Park - Tampa Bay Pilot Study.

BEAM TRAWL SAMPLE RESULTS							
	Sources			Controls			
Stations	4	6	5	7	2	1	3
Number of Taxa	8	13	16	29	27	31	38
Index Total	15	27	35	69	72	84	106
Index	1.88	2.08	2.19	2.38	2.67	2.71	2.79

13.4 North Carolina - *Comparison of Biological Metrics Derived From Ponar, Epibenthic Trawl, and Sweep Net Samples: A North Carolina Pilot Study*

13.4.1 Study Objectives

A test was designed to compare biological metrics derived from three sampling methods, to determine which methods and metrics best demonstrated differences between sampling sites (Eaton 1994).

Test data consisted of benthic assemblage collection results for petite ponar, epibenthic trawl and sweep net samples taken in the vicinity of Wilmington, North Carolina (Figure 13-1). The data set included February and May 1993 ponar samples, November 1993 trawl and sweep net samples, and February 1994 samples using each of the three collection methods.

13.4.2 Study Location

Three sampling sites were located in polyhaline (>20-ppt) waters in the Wilmington vicinity. Howe Creek, a primary nursery area north of Wilmington, was selected as a reference site. Development in the Howe Creek area was sparse residential on the north side of the creek with a new development on the south side. Samples were collected on the north side of the creek, which placed a large saltmarsh between the collection site and the development to reduce possible impacts. The Howe Creek sampling location was characterized by sand and shell substrata, abundant sponge and oyster populations, and seasonally abundant macroalgae (*Ectocarpus* and *Cladophora*).

A second sampling station (Hewletts Creek) was chosen as a test site for the assessment of nonpoint impacts. Hewletts Creek receives runoff from central Wilmington. It has occasionally received pump station overflows, and its shoreline is heavily developed with single family residences. The large quantities of macroalgae (*Enteromorpha*, *Ectocarpus*, and *Porphyra*) that have been flushed out of the creek could indicate potential excess nutrients. The Hewletts Creek station was characterized by hard-packed medium sand and shell substrates, intertidal oyster bars and saltmarsh.

A sampling station located in Bradley Creek was selected as a representative impaired area. Most of this watershed has been heavily developed, and the lower portion of the creek supports two marinas. Sampling was conducted just upstream from one large marina and immediately downstream from the U.S. Route 76 bridge. The Bradley Creek station was characterized by mud and muddy sand substrate, intertidal oyster bars, and seasonally common macroalgae (*Ectocarpus*).

13.4.3 Study Methods

Three types of gear were employed to sample the benthic assemblages at each station. A petite ponar was used to collect three replicates of 1-3 grabs each (depending on faunal density), thereby sampling the infauna in a 0.04-0.13-m^2 area at each station. An epibenthic trawl (1.25-m net mouth) was pulled over 4-m of unvegetated substrate to collect the epifauna and obligate infauna in a 5-m^2 area at each station. This method is further described in Section 9.5. A D-frame net was swept through all available habitats for 10-minutes, collecting the epifauna and shallow infauna in a 20- to 60-m^2 area. Advantages and disadvantages noted

for each collection method are listed in Table 13-10.

All samples were preserved in the field with 10% formalin with Rose Bengal dye added as a tissue stain. Samples were returned to the laboratory, where they were sorted from the detritus, then identified to the lowest practical taxonomic level (usually species).

Biological metrics taken from a wide variety of sources were tested for each sampling method. It was expected that different metrics would prove useful for different sampling methods. Test

Table 13-10. Advantages and disadvantages noted for the three benthic assemblage collection methods.

METHOD	ADVANTAGES	DISADVANTAGES
Petite Ponar	• can be used in any depth water on almost all substrates (except hard bottoms). • most previous researchers used dredges, therefore some comparisons with historic data can be made. • true replication allows for statistical treatment of the data.	• it samples a relatively small area, therefore rare and/or large taxa may not be collected. • the infauna are the most tolerant portion of the benthic community, therefore minor stresses may be easily missed. • sorting through large amounts of sediment and counting hundreds of individuals of one or two taxa can become tedious.
Epibenthic Trawl	• epifauna are generally more intolerant to stresses than infauna, therefore more subtle environmental changes can be detected than with infaunal sampling. • a larger area is sampled than with dredges, therefore more rare taxa should be collected. • when operated properly, a relatively small amount of sediment is collected, therefore sorting is not tedious.	• results are not comparable with most historic databases. • the trawl is fairly unwieldy and takes training to use properly. • it is impractical to use in depths beyond 5-10 m or in strong currents (>1.5-2 m/s).
Timed Sweep	• a large number of taxa are collected including rare, large and intolerant taxa. • since metrics are more reliable when calculated with increasing observations (taxa), change in a metric is a more reliable indicator of environmental change. • being semi-quantitative, only an estimate of abundance is required rather than having to count each individual. • all habitats are sampled, therefore loss or degradation of habitat is more readily documented.	• method is limited to wadeable areas. • large amounts of sediment are usually collected, making sorting tedious. • a higher degree of taxonomic expertise is required than needed for the other methods. • results are not comparable with most historic databases.

metrics included: Farrell Biotic Index (modified for North Carolina), number of amphipods and caridian shrimp, total taxa, percent annelid abundance, percent mollusc abundance, Shannon-Wiener diversity index, amphipod abundance, polychaete abundance, molluscan abundance, gastropod abundance, bivalve abundance, capitellid polychaete abundance, spionid polychaete abundance, Hurlbert's PIE, Keefe's TU, Simpson's D, and oligochaete and pelecypod abundance (Engle et al. 1994; Farrell, 1993b, Nelson 1990, Washington 1984, Weisberg et al. 1993). All metrics were tested using the data generated from each of the three collecting methods. A metric was deemed to work if it was able to correctly rank the stations; i.e., as reference, slightly impaired, or heavily impaired. Those metrics that correctly ranked the stations were further tested on a larger database to determine if metric ranking was a spurious coincidence or was due to the measurement of a consistent component of the biological community.

13.4.4 Results

Metrics that correctly ranked the three sites and their values, are listed in Table 13-11 by sampling method. The Biotic Index was the only metric to correctly rank the sites; i.e., as reference, slightly impaired, or impaired for each of the three collection methods. For samples collected by petite ponar, the Biotic Index correctly ranked the sites for the two February samples, but failed to correctly rank the sites in May. This may be related to seasonal fluctuations in recruitment (Holland 1985). The Biotic Index as a function of salinity at 38 sites sampled via petite ponar is presented in Figure 13-3. Diamonds represent reference sites, triangles represent impacted sites and squares represent areas of intermediate or unknown water quality. Lines in the figure represent possible break points for future criteria.

Two metrics, Biotic Index (BI) and % Oligochaete and Pelecypod abundance (% O&P), correctly ranked the three sites sampled using an epibenthic trawl (Table 13-11). In February, but not in May, the %O&P was low because two taxa made up 70% of the individuals at this site. This heavy skewness in abundance may be due to seasonal recruitment. To date, these samples are the only collections made using the modified trawl. More samples are required to adequately test the efficacy of the trawl.

The sweep method had three metrics that ranked the three sites correctly (Table 13-11): Biotic Index (BI), Total Taxa (TT), and Amphipod and Caridean Shrimp Taxa (A&C). Graphs of BI, TT, and A&C values for 63 timed sweep samples over a range of salinities are presented in Figure 13-4. Each metric appeared, in varying degrees, to be affected by salinity. At sites where salinities were above 8-ppt, there was sufficient separation between Reference sites (diamonds) and Impacted (triangles) sites to identify sites with Intermediate impact (squares) as well. This separation was smaller in intermediate salinities (8-20-ppt) than higher salinities (>20-ppt). Samples collected below 8-ppt salinity showed a limited range of metric values. Only BI was able to separate Reference from Impacted sites in these low salinities.

The Total Taxa metric may be related to the habitat diversity of an area; a diversity of habitats at a site would include more niches, thus allowing the survival of more taxa. This suggests that the Total Taxa metric could serve as a habitat quality measure as well as a measure of water quality.

Table 13-11. Functional metrics for the three benthic assemblage collection methods.

	Howe Creek (Reference Station)				Hewletts Creek (Nonpoint Source Impact Station)				Bradley Creek (Urban Impact Station)			
Petite Ponar												
Date	2/93	5/93	2/94	5/94	2/93	5/93	2/94	5/94	2/93	5/93	5/94	5/94
Biotic Index	2.7	2.1	2.1	1.9	2.0	1.6	2.2	1.9	1.6	1.9	1.8	1.4
Epibenthic Trawl												
Date	—	—	2/94 5/94		11/93	—	2/94	5/94	—	2/94	5/94	—
Biotic Index	—	—	2.5	2.8	2.8	—	2.4	2.7	—	1.7	2.0	—
% Oligochaeta & Pelecypoda Abundance	—	—	21	3	27	—	31	6	—	37	4	—
Timed Sweep												
Date	—	—	2/94	5/94	11/93	—	2/94	5/94	11/93	2/94	5/94	—
Biotic Index	—	—	2.4	2.5	2.2	—	2.2	2.0	1.8	1.8	1.9	—
Total taxa	—	—	109	95	94	—	91	105	45	60	68	—
Amphipoda & Caridean shrimp	—	—	17	22	9	—	9	15	7	7	9	—

Figure 13-3

Ponar samples: biotic index vs. salinity

Ponar Biotic Indices
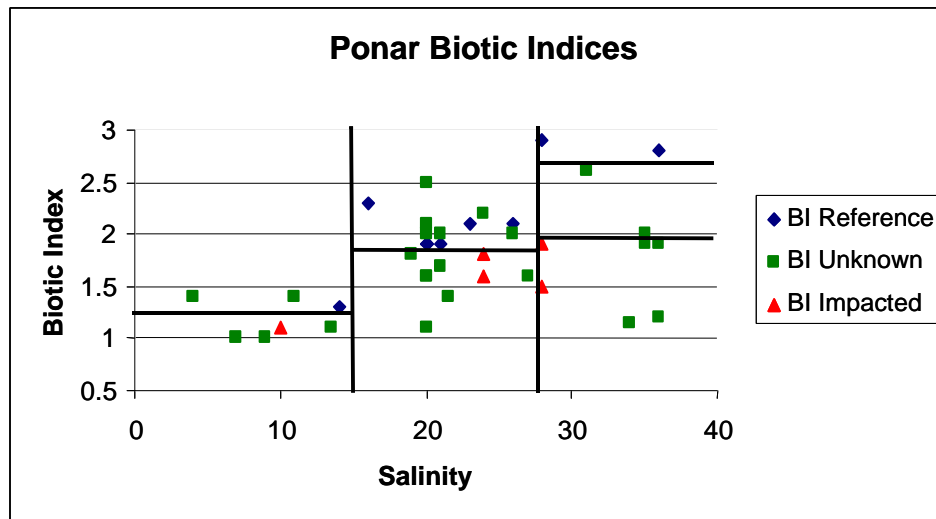
Figure 13-4
BI, total taxa and amphipod and caridean taxa by salinity.

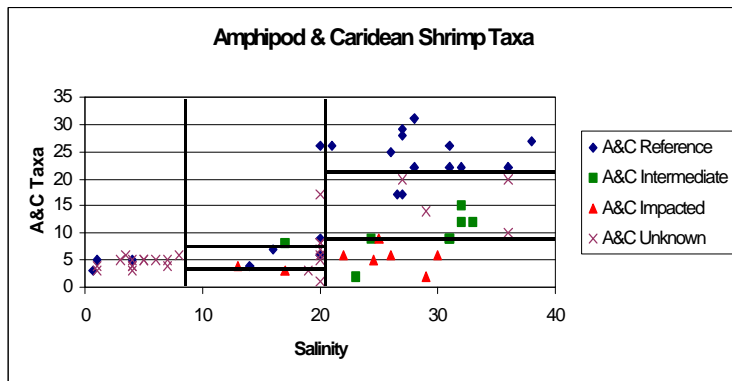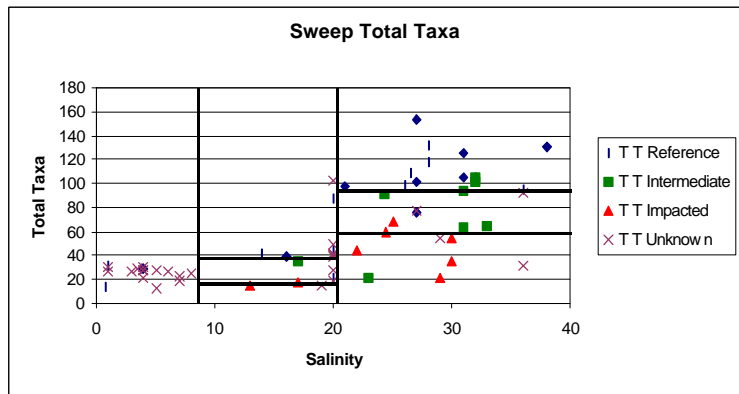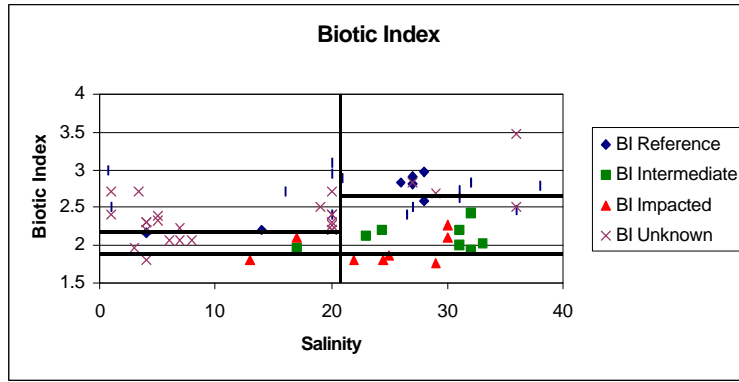Amphipods and caridean shrimp make up 10-15% of the total taxa at a site. This correlation explains why the graphs of the TT and A&C metrics look similar. Since the Crustacea include many of the most intolerant taxa in the estuary, the A&C metric may prove to be more sensitive to slight differences in water quality than the other metrics tested. One potential problem with the A&C metric is that it, like TT, appears to be affected by habitat quality, especially the presence or absence of seagrass and shells.

The next step, following method selection and metric determination, was biocriteria development. In this exercise, sweep samples at sites above 8-ppt were used because multiple metrics had been identified which showed a range of water qualities. For each metric, a value above the Reference/Intermediate line (Figure 13-4) was scored five points whereas a value below the Intermediate/Impacted line was scored 1. To increase sensitivity, the Intermediate Impact area was subdivided: values in the upper 20% were scored 4 points, values in the middle 60% were scored 3 points, and values in the lower 20% were scored 2 points. Points for each of the three metrics were summed, giving each site a total score between 3 and 15 points. Water quality bioclassifications were assigned based on the number of points scored by a site (Figure 13-5).

An attempt was made, in step three of biocriteria assignment, to address natural situations where Taxa Richness was depressed at a site (little habitat diversity, wide salinity swings, or high wave action). If one or more of these situations could be identified for a site, an extra two points were awarded to the total

score. While this appears to adequately correct a previously unaddressed problem in biocriteria development, assessment of the usefulness of this approach must await a validation study, which is beyond the scope of the exercise described here.

13.4.5 Summary

Test results indicated that there was no metric which consistently ranked the test stations in *a priori* order of impact based on petite ponar collections, though this may have been due to confounding by a spring peak in recruitment. The Biotic Index ranked sites correctly most often. Epibenthic trawl results correctly ranked the test sites using the Biotic Index and percent abundance of Oligochaeta and Pelecypoda metrics. Further sampling with the epibenthic trawl is required to determine whether it or the ponar will give more reliable results in non-wadable areas. The sweep method appeared to be the most versatile of the three test methods, resulting in three metrics that correctly ranked the test sites. All metrics appeared to lose sensitivity at salinities below 20-ppt. Possible seasonal effects and differences in substrate appeared to be confounding the analyses as well; therefore, these factors must be taken into account during the biocriteria development process. The Biotic Index appeared to be the most versatile tool since it was the only metric to correctly rank sites for all methods and all salinities. Initial efforts at biocriteria development in North Carolina will focus on the Biotic Index as well as on further sampling to determine the effects of seasonality, substrate, salinity, and habitat variables.

Figure 13-5

Development of biocriteria.

STEP 1: Assign points for each of three metrics from a sweep sample.

Polyhaline (21 ppt to seawater)

Points	5	4	3	2	1
BI	≥ 2.6	2.59 - 2.5	2.49 - 2.01	2.0 - 1.91	≤ 1.9
Total Taxa	≥ 95	94 - 86	85 - 69	68 - 60	< 60
Amphipods & Caridean Shrimp	≥ 21	20 - 18	17 - 13	12 - 10	9 - 0

Mesohaline (8 ppt to 20 ppt)

Points	5	4	3	2	1
BI	≥ 2.2	2.2 - 2.16	2.15 - 1.96	1.95 - 1.9	< 1.9
Total Taxa	≥ 38	37 - 32	31 - 24	23 - 18	17 - 0
Amphipods & Caridean Shrimp	≥ 8	7	6 - 5	4	3 - 0

STEP 2: Sum points. This will yield a number between 3 and 15.

STEP 3: Check for Bonus Point conditions. Add 2 points to score if one or more of the following conditions occurred: 1) Homogeneous habitat, 2) consistently high wave action, 3) very high (>26 ppt/yr) salinity fluctuations.

STEP 4: Assign Bioclassification.

Bioclassification	Points
No Impact	13-15
Slight Impact	11-12
Moderate Impact	8-10
Elevated Impact	6-7
Severe Impact	3-5

Primary Contact: Larry Eaton,
NC Division of Water Quality,
4401 Reedy Creek Road
Raleigh, NC 27602
919-733-6946
lawrence@dem.ehnr.state.nc.us

13.5 Indian River, Florida - *Field Verification of Marine Metrics Developed For Benthic Habitats: Indian River Lagoon, Florida Pilot Studies*

13.5.1 Study Objectives

A research program was developed for Florida estuaries to promote the identification of benthic marine parameters indicative of relative water resource quality (Nelson et al. 1993, Nelson and Spoon 1994 a, b). The development of these parameters or metrics was ultimately intended to help quantify the diverse attributes and interrelationships of the community to: enhance documentation of possible resource impairment from point and non-point sources; evaluate aquatic life use attainment; and to be incorporated in the biological criteria process.

13.5.2 Study Methods

The process of developing benthic biota community parameters (or metrics) indicative of Florida estuarine resource quality was initiated in 1993 with Indian River Lagoon pilot studies (Figure 13-1). The initial study involved the collection and analysis of samples taken from six stations, including two within the main Indian River Lagoon and four at the mouth of tributaries. Benthic samples were collected by a diver using 8.2-cm diameter Lexan cores which were sectioned into 0-5-cm and 5-15-cm depth fractions. Study sites were selected with the primary criterion of a presumed difference in pollution impact, with secondary emphasis on similarity of sediment type and tertiary emphasis on similarity in salinity. Three sites were designated impaired and three were designated low impairment sites. The small scale or smaller number of stations

limited the bottom salinity types sampled to mesohaline, polyhaline, and euhaline locations.

The nature of pollution impacts in the Indian River Lagoon presented a major problem in sample selection. Maximum impacts of pollution input (primarily from urban runoff) are felt within the small lagoonal tributaries as compared to the lagoon proper. Non-impaired tributary sites are generally not available, which forces most reference sites to be in the lagoon proper. This sometimes resulted in a difference in salinity between the impaired and non-impaired sites. For example, during winter sampling, mean salinity at impaired sites was 13-ppt (mesohaline), and was 25.3-ppt (euhaline) at non-impaired sites. However, these spatial salinity differences appeared to be seasonal in the lagoon. Samples taken in June had a mean bottom salinity of 25.3-ppt (euhaline) at impaired sites compared to 29.6-ppt (euhaline) at non-impaired sites.

13.5.3 Study Results

The benthic data were summarized in terms of: the mean percent of biomass contained in the top five centimeters of the sediment profile at each of the stations; and the mean weight per individual compared among sites and between the 0-5-cm and 5-15-cm depth segments. There was no clear difference in mean percent biomass (contained in the top 5-cm) between the two sets; i.e., impaired, non-impaired of sites. There was also no clear difference in mean weight per individual based on presumed differences in pollution impact either for the surface sediments or for the deeper sediments. Biomass differences with differing salinities also showed no clear differences, in that mean values of the percent total biomass above 5-cm ranged from 68% to 89% in

mesohaline areas, and 69% to 81% in euhaline areas. The value in the single polyhaline area was 94%.

Sediment types within the study area were classified as sand (>70% sand), mixed (30-70% sand), and mud (<30% sand). Both the impaired and non-impaired sites had all sediment types represented. There were no apparent trends in biomass data among the sediment types. There was no indication that sand sites had less biomass in surface sediments than mixed sediments, or that mean weight per individual differed among sediment types.

During the initial studies, benthic data were also summarized in terms of total individual and total species metrics. The mean percentage of total individuals present above 5-cm ranged from 96 to 99.6% at the study sites. The differences in this metric between sites of different pollution impact was thus very low; therefore, this metric did not clearly distinguish Indian River Lagoon sites. Mean percentage of species above 5-cm in the sediments was calculated from the data by dividing the total number of species in the 0-5-cm fraction by the sum of this value plus the total number of species recorded in the 5 -15-cm fraction for each site. There was no clear separation between the sites based on this metric.

The initial Indian River Lagoon pilot study and previous studies of the area have examined the following metrics:

▸ Mean total number of individuals;

▸ Mean total number of species;

▸ Percentage of amphipods;

▸ Percentage of spionids;

▸ Spionidae/capitellidae ratio;

▸ Apparent color RPD depth.

Of the seven metrics, separation between impaired and low impairment sites was good for mean number of species, percentage of amphipods, percentage of spionids, spionidae/capitellidae ratio, and apparent color RPD depth. The percentage amphipod, percentage spionid, and spionidae/capitellidae ratio metrics require separation of individual specimens which requires greater time than simple counts of total individuals or total biomass. However, these metrics seemed to offer much greater powers of resolution than measures of total individuals or biomass.

Limitations on the generality of the conclusions of the initial pilot study were imposed by the limited number of sampling sites (6) and by the fact that samples were obtained at only one point in time. Seasonal variation in benthic systems can be substantial; therefore, it was essential to verify the temporal generality of initial conclusions. Similarly, spatial variations in salinity regime have been demonstrated to influence metric values. Therefore, more extensive spatial and temporal sampling was warranted to verify the utility of the proposed metrics. To provide temporal verification of metrics, the six sites originally sampled in January 1993 were resampled, and two new sites were added to the sampling plan to represent additional spatial coverage of the Indian River Lagoon. The two additional sites were located near Cocoa, Florida, with one presumed to be an impaired; i.e., located near a sewage outfall pipe, and the other a low impairment site. Core samples were collected by divers (as in the initial pilot study) during June, July and August 1993. All organisms collected were

subsequently identified to the lowest feasible taxon, counted, oven dried, and weighed in the laboratory. Surface water temperatures and salinity, and bottom salinity measurements were made in the field, and sediment samples were collected by skin diving. One of the proposed metrics was to involve visual determination of the apparent color RPD depth. It was difficult to measure a visual RPD with any degree of confidence since the surface layers of sediment were often flocculent and were therefore disturbed by the coring process. Attempts to measure the apparent visual RPD were therefore abandoned during the second phase of pilot studies.

A total of 64 benthic cores and 128 core fractions were collected and processed during the second phase of pilot studies. There was no clear distinction for the 0-5-cm sediment fraction in mean number of individuals per core recorded at low impaired versus impaired stations. Mean number of benthic taxa recorded per core from the 0-5-cm fraction also failed to show clear differences between the two sets of stations. Clear distinctions were observed, however, between the impaired and low impaired sites with respect to mean abundance per core for benthic organisms in the 5-15-cm fractions. Abundances in the 5-15-cm fractions differed by a factor of at least 4 between the two sets of stations; i.e., mean equaled 0-1.3 at impact stations and 5-10 at low impact stations. Impaired and low impaired sites also showed clear separation based on the mean number of taxa per core in the 5 - 15-cm fraction.

The total species richness of amphipod crustaceans was seven at both the impaired and low impact sites, and the species recorded were similar. The metric based on the percentage abundance of amphipod crustaceans;

i.e., percent of total, failed to clearly distinguish the two types of sites. Comparisons of amphipod total abundance; i.e., amphipod total count, also failed to clearly distinguish impaired versus low impact sites, ranging from 4-1789 at low impact sites and 27-62 at impaired sites. The 1993 summer results contrast strongly with the winter results for the amphipod metrics. Winter data showed clear separation of impact versus low impact stations with respect to both percent abundance and total number of amphipods, whereas summer data did not.

The ratio of spionid polychaete abundance to capitellid polychaete abundance showed only partial separation between station types in the summer samples. There was a decreased degree of separation with this metric in summer versus winter samples. The differences in the ratio were generated both by reduced numbers of spionids and by increased numbers of capitellid polychaetes at the low impact sites for winter samples. In contrast, differences in summer samples for the low impact sites were mainly caused by high values for capitellids. Therefore, a total capitellid metric was examined for both seasons. Clear separation between impaired and low impact sites was given for this metric for both summer and winter data. Examination of a total annelid abundance metric also demonstrated separation of impaired and low impact sites (for summer data).

Total faunal biomass showed no separation of stations for either the 0-5-cm or 5-15-cm core fractions. Expression of the biomass values above 5-cm were modified by subtracting biomass for the occurrence of a few large organisms (e.g., large bivalves). The adjusted surface biomass metric also failed to

clearly separate the impaired and low impact sites.

Differences in the values of benthic community parameters were apparent in summer samples as compared to the winter samples from the same study sites. The seasonal changes in abundance were anticipated given previous knowledge of seasonal abundance patterns of macrobenthos in the Indian River Lagoon. Some proposed metrics were consistent in their performance in both winter and summer samples (Table 13-12). Both abundance and taxa richness in the deep sediment fraction were metrics which gave clear separation in the sets of stations in both winter and summer. Abundance of capitellids also consistently separated the station types during both seasons.

The performance of some of the metrics which appeared promising in the winter samples was somewhat altered in summer. For example, taxa richness in the 0-5-cm fraction, percent amphipod abundance, total amphipod abundance, and spionid/capitellid ratio metrics discriminant stations in the winter, but did not do so (or gave unclear results) in the summer. Explanations for this change in performance may be complex.

The biomass measurements used in the pilot studies were made on specimens separated into lowest identified taxonomic units, which required considerable time and effort. Had the biomass measures provided clear separation of station types, it would have been warranted to suggest that all specimens be pooled to obtain a single biomass value. However, it did not appear that biomass values for either depth fraction were useful as a benthic metric for the Indian River Lagoon.

The pilot study results clearly indicate that the season during which sampling takes place may influence the ability of a given metric to distinguish among sites. Overall, clearer separation was seen among sets of stations for winter sampling than for summer sampling. This appeared to be related to the fact that highest organism density in Indian River Lagoon benthos is seen during late winter, rather than in the summer as is the case at other locales. This clearly points out the need to evaluate seasonality at specific geographic areas.

Relatively few of the proposed metrics consistently separated sites in the Indian River Lagoon. The mean abundance of organisms and mean species richness in the 5-15-cm depth fraction, and capitellid abundance metrics all provided consistent separation of station types. The relatively small sample size in terms of number of stations appeared to result in ambiguous interpretation; i.e., clear station separation ability in winter and marginal in summer, for the total amphipod abundance and spionid/capitellid ratio metrics. The natural temporal variability in the benthos may be sufficiently extreme to affect the performance of these metrics; therefore, the best way to minimize the influence of the variation may be to sample as many stations as possible.

In the most recent phase of pilot studies, two amphipod metrics - mean number of amphipods per site and the ratio of Corophiidae/(Ampeliscidae + Phoxocephalidae) - were assessed at a total of ten stations within the Indian River Lagoon. The original eight pilot sites were resampled and two additional sites were sampled during May and June 1994, using techniques as described for the earlier pilot studies. A total of 80 benthic cores were collected and processed.

Table 13-12. Comparison between winter and summer samples of the ability of the various metrics tested to discriminate between impaired and low impairment sites.

METRIC	WINTER	SUMMER
0-5 cm abundance	NO	NO
5-15 cm abundance	YES	YES
0-5 cm taxa richness	YES	NO
5-15 cm taxa richness	YES	YES
percentage amphipods	YES	NO
total amphipod abundance	YES	?
spionid/capitellid ratio	YES	?
capitellid abundance	YES	YES
total annelid abundance	NO	YES
total biomass	NO	NO
total biomass (excluding large bivalves)		?
mean percent biomass above 5 cm	NO	NO

NOTE: ? indicates marginal utility of metric due to inconsistent discrimination of impaired and low impairment sites.

Results of collection analyses showed that the simplest amphipod metric, mean total abundance, clearly separated impaired from low impact sites in the late winter samples taken in 1993. However, summer 1993 and 1994 results indicated that the response of this metric was not satisfactory. Available water quality information suggests a division of the set of 10 stations into three groups: high impact, moderate impact, and low impact. Use of the mean number of amphipod metric did not provide a similar separation of sites for summer 1994 sampling data. However, the outcome of the Corophiidae/ Ampeliscidae metric calculations was most consistent with the high impact, moderate impact, low impact division of sites, and therefore appeared to reasonably reflect water quality conditions of the Indian River Lagoon.

Primary Contact: Dr. Walter G. Nelson, National Health & Environmental Effects, Research Laboratory/ORD Western Ecology Division, Hatfield Marine Science Center 2111 SE Marine Science Dr. Newport, OR 97365-5260 541-867-4041 Nelson.Walter@epa.gov

13.6 Ocean City, MD — Bethany Beach, DE — A Preliminary Study of the Use of Marine Biocriteria Survey Techniques to Evaluate the Effects of Ocean Sewage Outfalls in the Mid-Atlantic Bight

13.6.1 Study Objectives

This project investigates the practical, low cost application of marine biological community measurements and the near field/far field survey technique for use by coastal States as a water resource quality management tool. The methods applied here are derived from work reported by Pearson and Rosenberg (1978) and Mearns and Word (1982) with modifications.

13.6.2 Study Methods

The study area is a 16-km coastal reach between Bethany Beach, Delaware and Ocean City, Maryland (Figure 13-1, 13-6). These are nearly adjacent resort communities on the Mid-Atlantic seaboard between Delaware Bay and Chesapeake Bay. Each has a secondary treatment municipal sewage discharge site about 2.8-km offshore. Discharge is in both cases through a diffuser at a water depth of approximately 12-m. The Bethany Beach sewage treatment plant average discharges about 0.61-m^3s^{-1}(14-mgd) and Ocean City about 1.4-m^3s^{-1} (32-mgd).

A series of nine north-south trending stations were installed parallel to the coast at intervals of about 2-km, each in about 12-m depth of water and over medium to fine sandy bottoms to obtain a similarity of habitat as much as possible. The stations are labeled "A" through "I", with station "C" at the Bethany Beach outfall and station "G" at the Ocean City outfall (Figure 13-7). This structure provides a set of control or reference stations for comparison to the test stations at "C" and "G". Each station is located with differential GPS with an estimated precision for the receiver of +/- 5-m.

The variables measured are benthic fish and macroinvertebrate communities as reflected in indexes and metrics incorporating number of taxa and number of individuals per taxa. Fish surveys are made with a 6-m (5-m effective opening), 2.5-cm mesh otter trawl. Tows are made parallel to the shoreline at 1-ms^{-1} over 0.9-km with the station coordinates located at the mid-point of the tow. Trawl scope used is six to one. Benthic macroinvertebrate samples are collected with a 0.1-m^2 Smith-McIntyre grab or with a 0.1-m^2 Young grab, and three replicates are taken for each sample at each station site as indicated by DGPS coordinates. Ferraro et al. (1994) reviewed their extensive data base and concluded that five replicates with a 0.02-m^2 petite ponar grab, each sub-sampled with four 8-cm diameter cores is optimal for waters of the Southern California Bight. We elected to use the 0.1-m^2 grab with three replicates, but to count the entire grab. This was judged to be a reasonable compromise between more replicates and the uncertainty of sub-sampling a site for which there was inadequate preliminary information. From this data base we hope to make further sampling refinements in the future. Identifications of collected organisms are to species whenever possible. All survey work was conducted from the USEPA Ocean Survey Vessel Peter W. Anderson. The Anderson is a 50-m research ship, but all equipment used and methods employed are appropriate for deployment from a 15-m vessel typically used by most coastal States. Incidental to this project,

Figure 13-6

Bethany Beach - Ocean City study area.
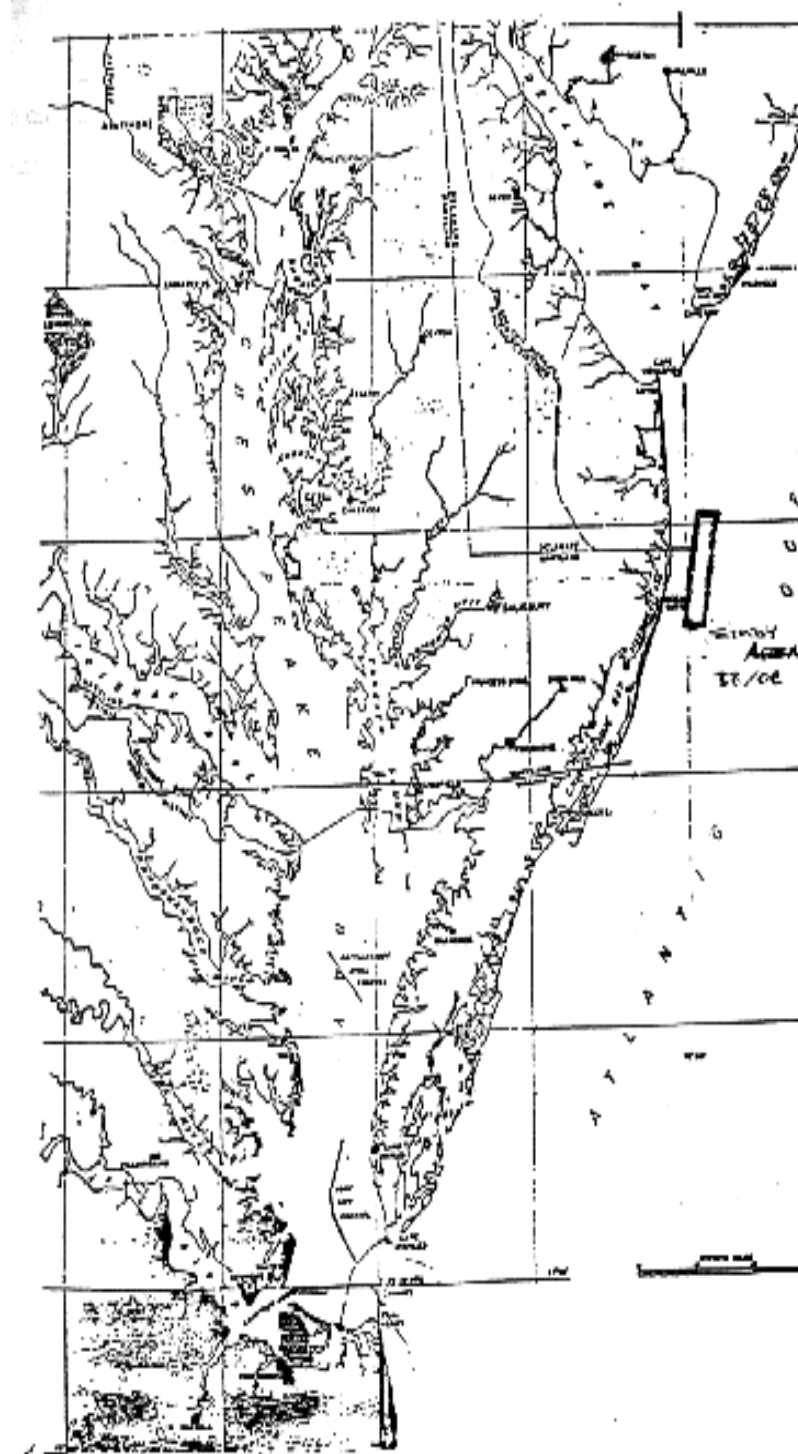
a comparison of the Smith-McIntyre and Young grabs was made. Six replicate drops with each grab were made at three sites characterized by: hard packed fine sand; medium grain sand; and coarse sand and gravel. Either gear was judged acceptable, but the Young grab was less inclined to wash.

Sampling surveys have been conducted twice a year in July-September and January-February since 1993 to determine if multiple season indexing is necessary or appropriate. While the Mid-Atlantic area is considered to have four discrete seasons, benthic communities are expected to be in flux

Figure 13-7

Bethany Beach - Ocean City sampling locations.

7.2 NATIONAL BIOLOGICAL CRITERIA STUDY AREA

(OCEAN CITY/BETHANY BEACH)

A

B

C   Bethany Beach Ocean Outfall

D

DELAWARE

E

MARYLAND

F

G   Ocean City Ocean Outfall

H

I

● - SAMPLING COORDINATES
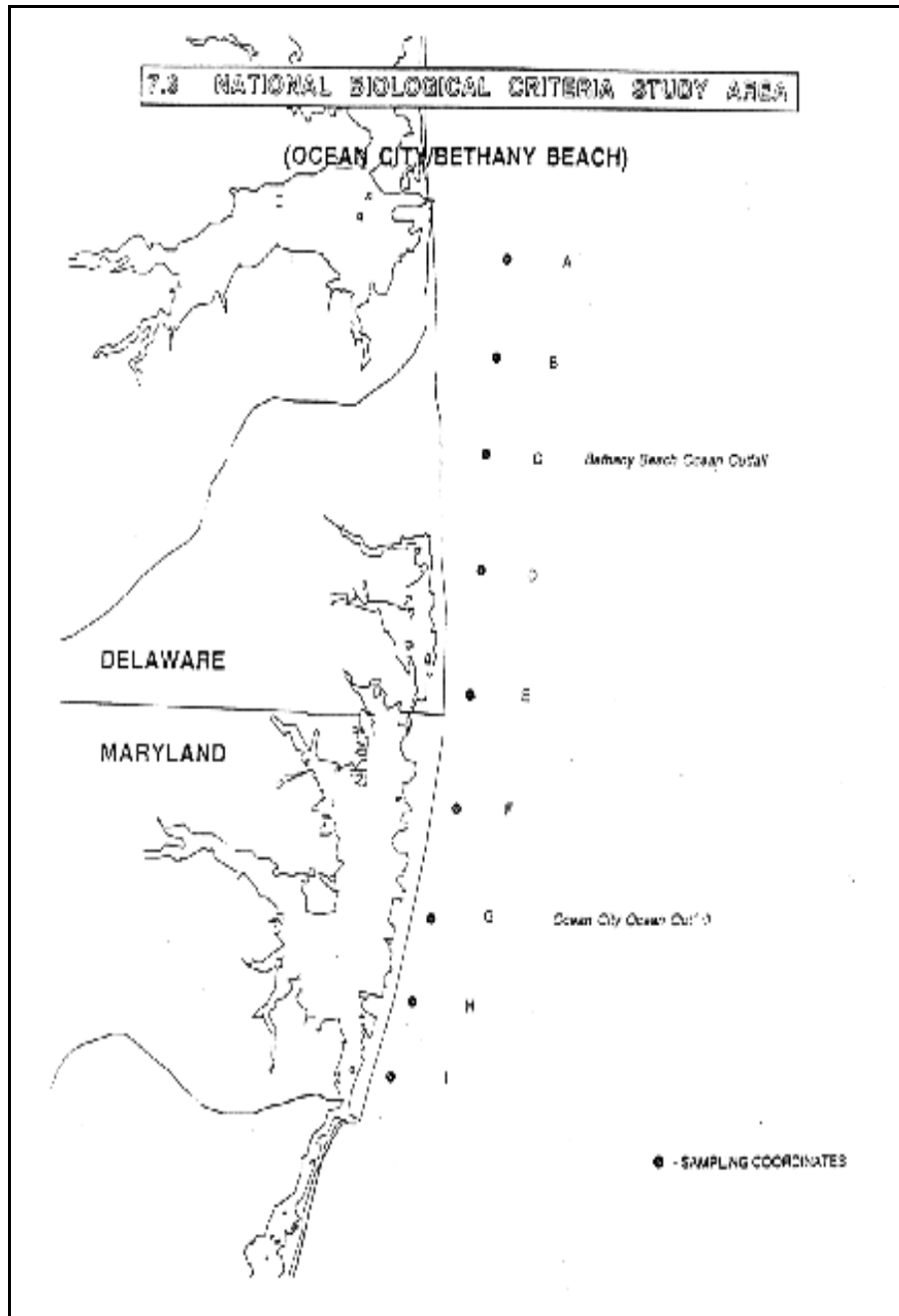
during Spring and Fall and to be most stable in Summer and Winter (Ranasinghe et al. 1994).

Fish sample processing is conducted on board with all individuals identified to genus and usually to species. Length measurements (TL) are made and any gross anomalies recorded. The fish are returned to the water as soon as measurements are completed. Benthic invertebrate sediment samples are sieved on board using a 0.5-mm mesh screen after recording a physical description of the sample and taking a 2.5-cm diameter subcore for grain size analysis. The retained material is fixed in 10% buffered formaldehyde with Rose Bengal dye added. Taxonomic identifications and counts are made later at laboratory facilities ashore with most identifications carried to the species level.

To make comparisons between the sample sites, habitat control in the survey design was maintained as well as possible by attention to three major variables:

1) Grain size of bottom sediments. This also reflected a habitat impact of the discharges when fine sediments were deposited in bottom depressions near the outfalls. At the beginning of the project, sediment samples were collected from all nine stations and analyzed for heavy metals and a for a standard array of toxicants. All results were insignificant, suggesting no other sources of biotoxicity or impairment indigenous to the immediate area.

2) Water depth. Water depth over stations "A" through "E" ranges from 11-m to 14-m with the variation accounted for by a general ridge and swale bathymetry off Bethany Beach. From "E" through "I", the variation is from 14- to 16-m, accounted for by an east-west ledge with about a 3-m drop just south of the Ocean City outfall. Subsequent data analyses suggest that these variations in water depth do not restrict fish or invertebrate distributions over the area.

3) Water quality. At the outset of the study, and each time biosurveys are conducted, multiple depth and standard water quality measurements are made using a Sea-Bird SBE-9 "CTD" probe. Conductivity, temperature, depth, dissolved oxygen, pH, transmissivity, and chlorinity/salinity are measured

and recorded throughout the water column.

To date, these variables have been consistent over the length of the transect for each cruise.

In keeping with the objective of low cost, practical applications of biological community measurements for resource impact detection; standard, basic but robust taxonomic indexes were applied to the data. The underlying premise for the indexes is that once the raw data for species and numbers of individuals per species are compiled, the investigator's primary question is whether or not there is a detectible impact. More refined indexes and indicators can later be applied or developed as needed. In this regard, the treatments selected for this project were: total number of individuals, total number of taxa (species), evenness index, Simpson's dominance index, Margalef's taxa richness index, and Shannon-Wiener index of general diversity. The appropriate equations were taken from Odum (1971).

13.6.3 Study Results

Fish Survey Data

Analysis of the fish data showed no significant differences in trawl data between the stations in either summer or winter collections for either number of taxa or numbers of individuals. These results are based on single tows at each station twice a year (summer and winter) for three years. Concern that this response results from too little data led to a trial in summer 1995, with three replicate trawl surveys over the nine stations, i.e., sequential tows of stations "A" through "I" conducted three times in one day. The results were still insignificant. Better results might be possible by replicating each station

individually. Eaton (1994) reported that for West Coast fish surveys in a Washington estuary, four replicate tows per station are necessary to obtain meaningful data within a fairly confined waterway. Qualitatively, taxa and number of individuals overall shifted considerably between summer and winter surveys at the nine stations. Greater numbers of both species and individuals (excepting winter runs of striped anchovy, *Anchoa hepsetus*) occur in the summer surveys.

Benthic Macroinvertebrate Data

Benthic macroinvertebrate results have been much more promising, but the same seasonal trend observed for fish for number of taxa and number of individuals has prevailed. Summer measurements are much more indicative of the condition of the benthic macroinvertebrate assemblages (some of the winter data is incomplete). The data in this instance is for three replicates at each station twice a year for three years. Significant differences are evident between each of the outfall sites and the other stations in the summer data (Figure 13-8a,b). The graphic data for number of individuals is intriguing in that it suggests enhanced and or enriched conditions at station "A", perhaps from the Delaware Bay discharge, and at the Ocean City outfall site.

When numbers of species are compared, a more negative trend in outfall impact is evident, especially for the Bethany Beach outfall station (Figure 13-9a). A similar pattern occurs at Ocean City, but is not as strong (Figure 13-9b). Ludwig and Reynolds (1988) state that a simple count of the number of species present, for samples of equal size, avoids some of the problems of using indexes which combine and may confound a number of

variables that characterize community structure.

However, in this instance, it appears that at least some indexes enhance the measurement of outfall perturbations. Box plots of Simpson's dominance index (Figure 13-10a, b), the Shannon-Wiener index of general diversity (Figure 13-11a, b), and particularly Margalef's richness index (Figure 13-12a, b) (Odum 1971), over the three years of summer data provide strong indications of the negative effect of both discharges on the benthic macroinvertebrate community.

13.6.4 Discussion and Conclusions

The nearfield/farfield survey design for biological surveys, together with basic indexes of community structure, appears to work equally well on the West coast and in Mid-Atlantic coast open water environments (Santangelo, pers. comm. 1996). If the investigator is careful to control for habitat characteristics, the ends of the transect can serve as a reference condition, the outfall stations as test sites, and the intermediate stations provide an indication of the gradation of impact(s). The nine station design of this study made it possible to treat the data as a combination of two impact sites on the ambient environment, or as two individual studies in tandem.

Summer benthic macroinvertebrate data from stations "A" and "C" were significantly different in either case, lending confidence to the conclusion that the wastewater discharges were having a measurable impact on the coastal marine environment. This is of particular interest because routine water quality and sediment investigations at the sites failed to consistently detect change between the outfalls and the surrounding stations. The biocriteria technique employed appears to be not
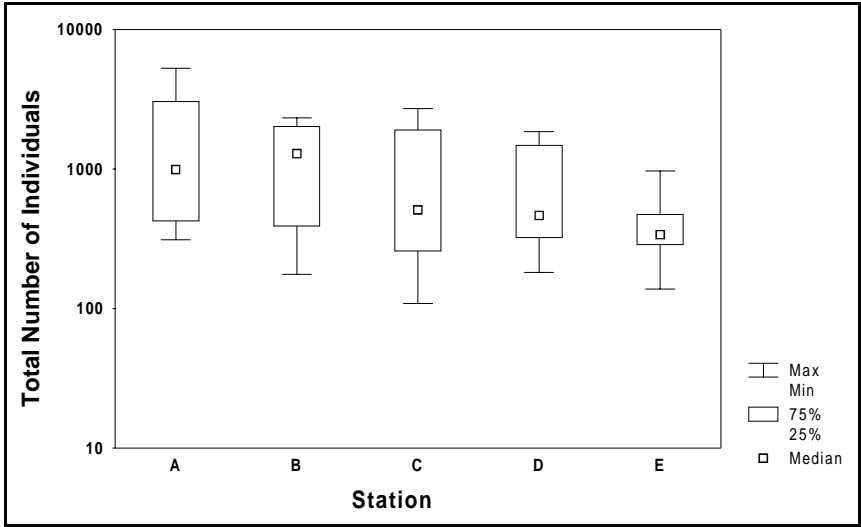
Figure 13-8a

Total number of macro-invertebrate individuals at Bethany Beach sites; summer data, n=9.
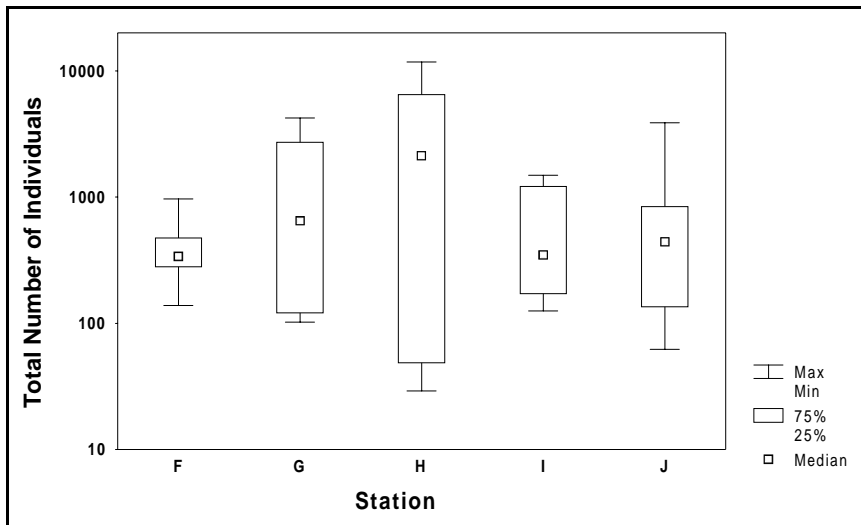
Figure 13-8b

Total number of macro-invertebrate individuals at Ocean City sites; summer data, n=9.

Figure 13-9a

Total number of macro-invertebrate taxa at Bethany Beach sites; summer data, n=9.
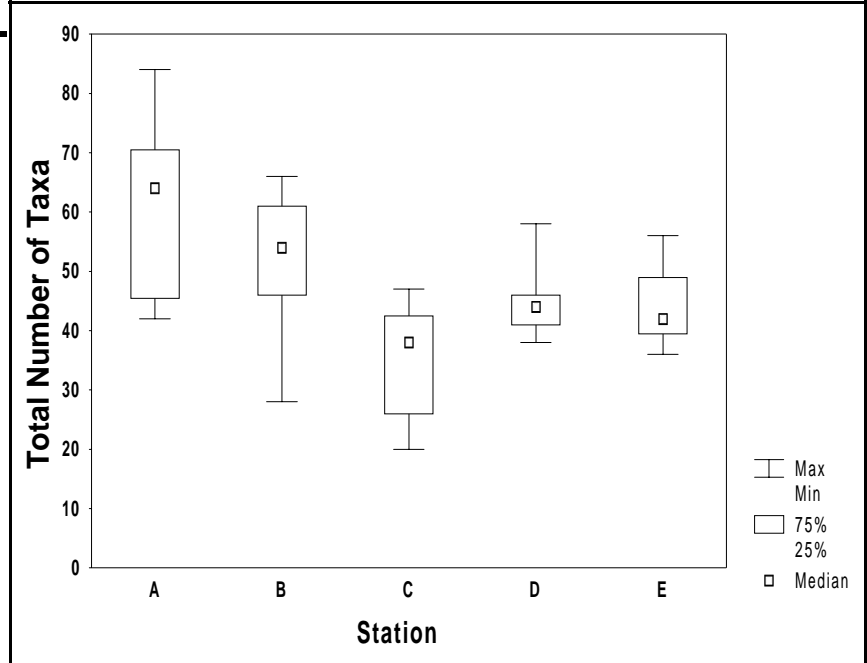
Figure 13-9b

Total number of macro-invertebrate taxa at Ocean City sites; summer data, n=9.
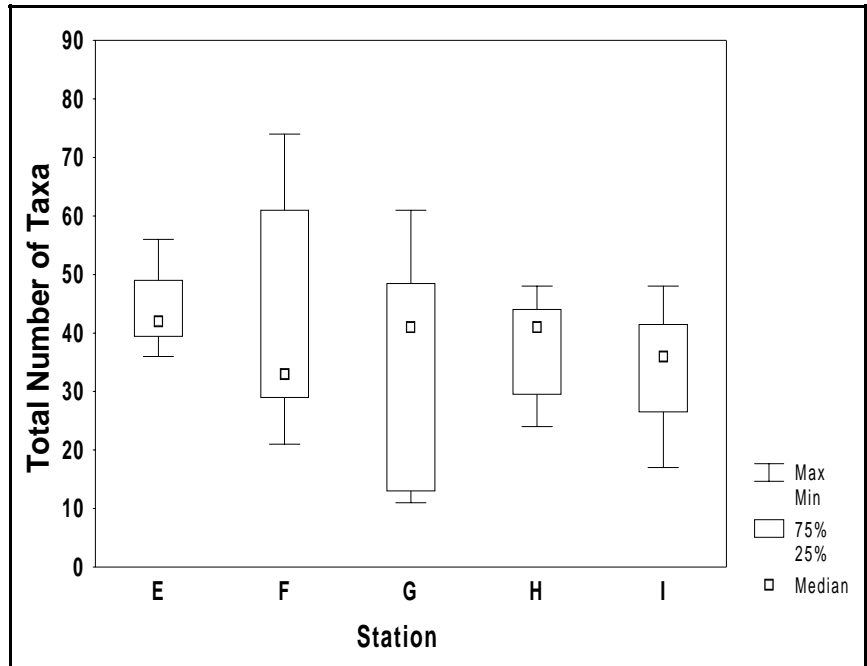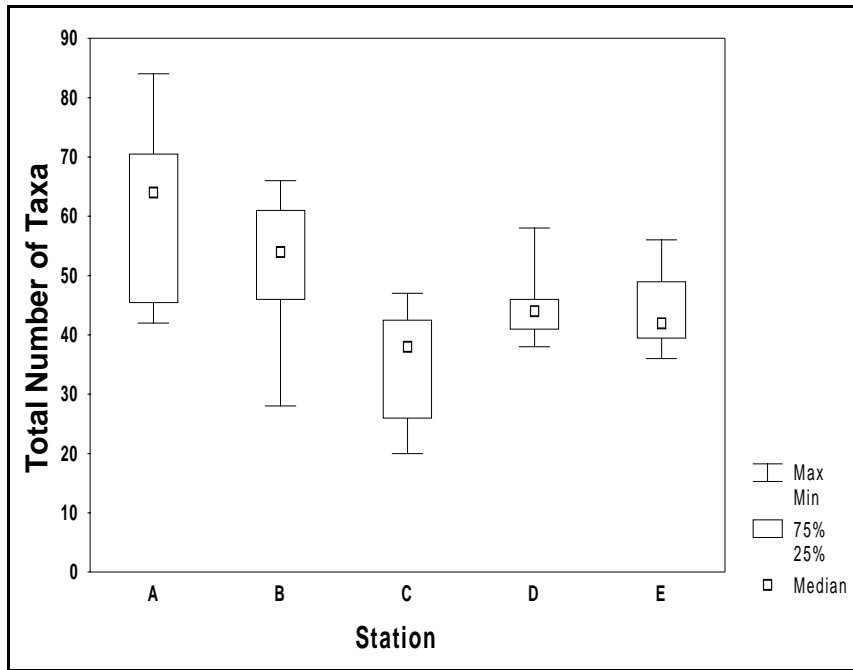
Figure 13-10a
Simpson's dominance index for macroinvertebrates at Bethany Beach sites; summer data, n=9.

Figure 13-10b
Simpson's dominance index for macroinvertebrates at Ocean City sites; summer data, n=9.

Figure 13-11a

Shannon-Wiener diversity index for macro-invertebrates at Bethany Beach sites; summer data, n=9.

Figure 13-11b

Shannon-Wiener diversity index for macro-invertebrates at Ocean City sites; summer data, n=9.

Figure 13-12a
Richness index for macro-invertebrates at Bethany Beach sites; summer data, n=9.

Figure 13-12b
Richness index for macro-invertebrates at Ocean City sites; summer data, n=9.
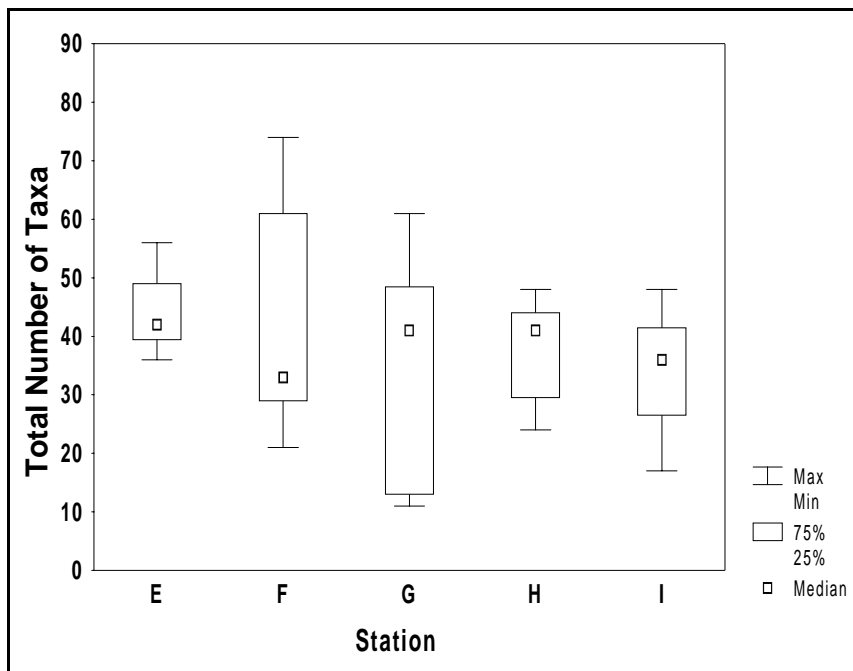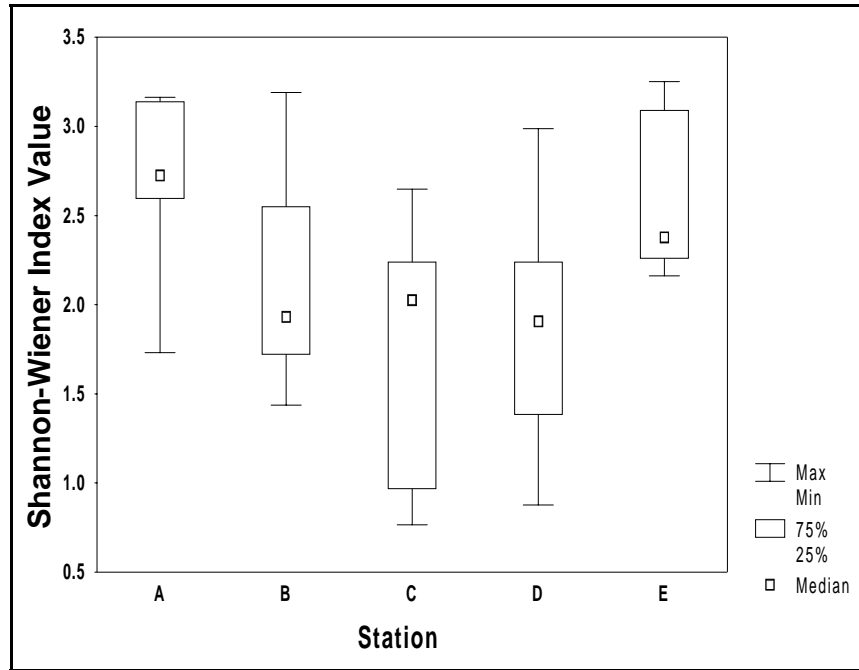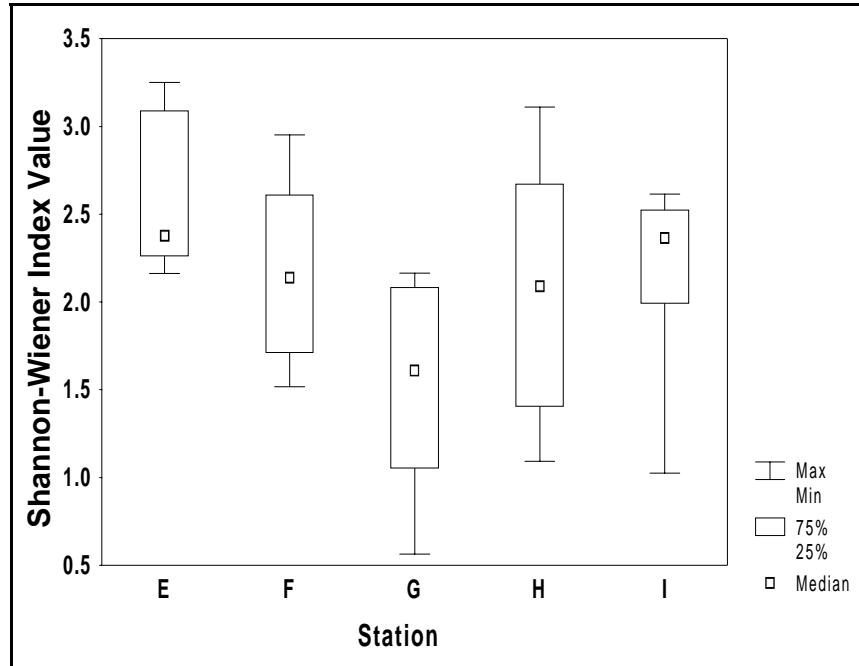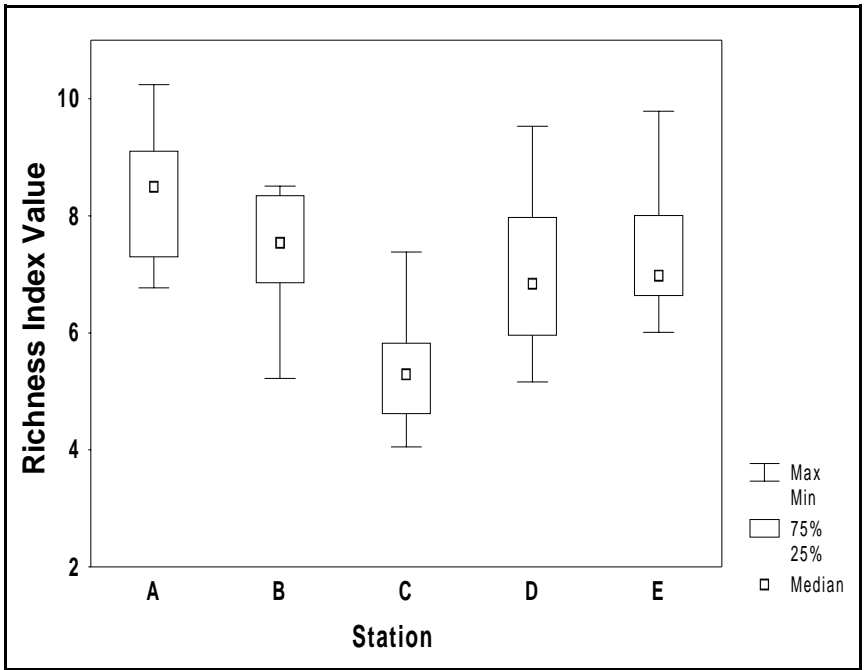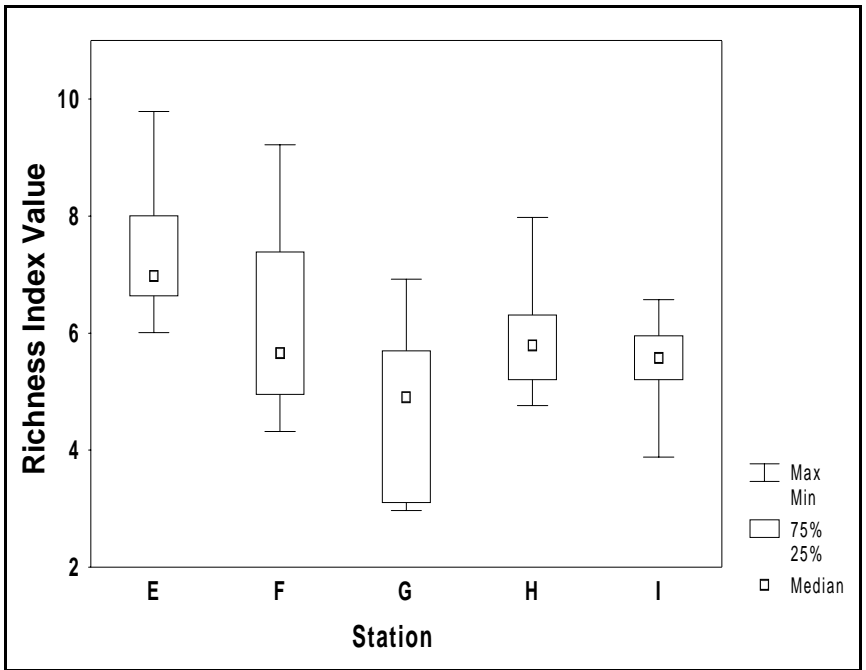
only practical, but sensitive as well, detecting impacts that might not always be observed with routine chemical testing. Standard indexes such as Margalef's Richness Index, Simpson's Dominance Index and Shannon-Wiener's Diversity Index are robust and were entirely appropriate for this survey.

The Smith-McIntyre and Young grabs were both entirely adequate, but the Young was more efficient and safer to work with, while the Smith-McIntyre was more accessible through the top for sub-sampling. Three replicate grabs were sufficient to generate meaningful data, but it may be possible to reduce the costs of the taxonomic operation by sub-sampling the grabs. An attempt was made to do this by mechanically splitting the intact samples in half with a sheet metal partition. It failed because the surficial organisms were unequally distributed as the sample drained and the ship rolled. Subcores of 5-cm diameter might be a better alternative requiring far less analytical effort. Similarly, sieving and counting only the top 5-cm of the sample as a variation of the technique reported by Diaz in Gibson et al. (1993) might be a more cost-effective approach. Another alternative to reduce the number of organisms dealt with is to double the sieve size to 1.0-mm, as practiced by many investigators. Any of these options could be explored and adopted as a cost-effective way to accomplish the benthic macroinvertebrate counts as long as the investigator ascertains that they produce reliable results consistent with those derived from the larger grab samples.

The 6-m otter trawl used in the fish surveys performed well and is believed to be appropriate for both coastal and estuarine biosurveys. However, the fish community does not appear to be very responsive to sewage discharge effects in this coastal area. This is probably because of the mobility of the fish in these open coastal waters, their seasonal migrations, and the potential sport and commercial fishing pressure confounding the survey effort; but the sampling replication factor was not adequately investigated in this study.

For biocriteria development and site monitoring, it is important to account for seasonality. For the Mid-Atlantic Bight, late June to early September appears to be a time of relatively high, stable community productivity and an optimal index period if once a year sampling is preferred. According to the Delaware and Maryland chambers of commerce, since Bethany Beach and Ocean City are summer resort communities, their populations increase at least ten-fold in warm weather (pers. comm. 1990). Their lower winter discharge rates, together with a natural cyclic depletion of the marine community, may account for the failure of our data to reveal sewage impacts in this season. This may not be the case with a year-round municipality of fairly large size. In any case, if the responsible agency can afford to sample at least occasionally in winter, that baseline biological data may prove invaluable in the event of oil spills or other marine accidents.

After the assessment of results from an initial set of 1.6-km interval station transects, the investigator may choose to delete some of the intermediate reference stations and replace them with a more diagnostic set of near discharge monitoring stations. It will then be possible to assess the relative expansion or contraction of the area of impact over time and in response to

plant operation declines or improvements. The initial and subsequent reference site data become part of the biocriteria which can be used as a benchmark to assess operational efficiencies, management initiatives, and the adequacy of NPDES permits.

Additional investigations for which results are pending

1. Because vertical splitting of each grab sample was determined to be an unsatisfactory approach to reducing sample volume and cost, we are attempting to test a horizontal approach at approximately the 5-cm depth level because most of the organisms observed are predominantly surficial sediment dwellers. In September 1996, the stations were sampled with three replicate grabs as before, but approximately the top 5-cm of sediment was scraped off of each sample and sieved through a 0.5-mm mesh screen. The remainder of the sample was similarly processed. We will count both fractions, combine the results and evaluate as usual.

This information will then be compared to a similar assessment using just the top 5-cm fraction. If the same impact information results, it may be possible to monitor the stations using just the surface fractions as long as these results are periodically calibrated against full grab counts.

2. On the January, 1997 survey, all of the stations sampled for benthic macroinvertebrates were sieved first through a 1.0-mm screen and then through the 0.5-mm screen. These separate fractions can be combined to produce a comprehensive result. The 1.0-mm fraction can then be compared

to this control to assess the relative efficacy of this technique as a cost saving approach for these waters. The process was repeated during the summer of 1997 and the results of both trials will be evaluated when taxonomic studies, which were delayed (for this and the above study) are completed.

3. Because of the promising results of this project so far, three additional stations have been added around each outfall station, e.g., "C" at Bethany Beach and "G" at Ocean City. The pattern creates a roughly equilateral triangle with approximately 0.46-km legs and a station at each apex with the original station in the middle of the triangle. The intent here is to see if it is possible to refine the spatial assessment of the zone of impact for each outfall analogous to the concept illustrated in Figure 13-13.

13.6.5 Use of the Bethany Beach-Ocean City Data to Illustrate Biocriteria Development

An example of biocriteria development using this pilot project is as follows.

Classification and Reference Site Selection: A review of the data as presented in Figures 13-7 through 13-12 suggests that stations A, E, and I are appropriate reference sites being at the center and extreme ends of the transect and equidistant from the defined locales of effluent discharges being evaluated. General water quality conditions, including salinity and depth, are consistent for all stations. Grain size, although shifting from sand and gravel in the north at station A to sand at station I in the south represents the general benthic habitat condition of the area with an

Figure 13-13

Proposed diagnostic nearfield station array.

net longshore current

• A
• B
• C
D1 • • D3
discharge • D • D4
D2 • • D5
• E
• F
• G

Data can be used to plot not only impact, but relative expansion or contraction of the area of impact over time.

#

A   B   C └─────┘ E   F   G
            D
         SERIES
    ◄─── STATIONS ───►
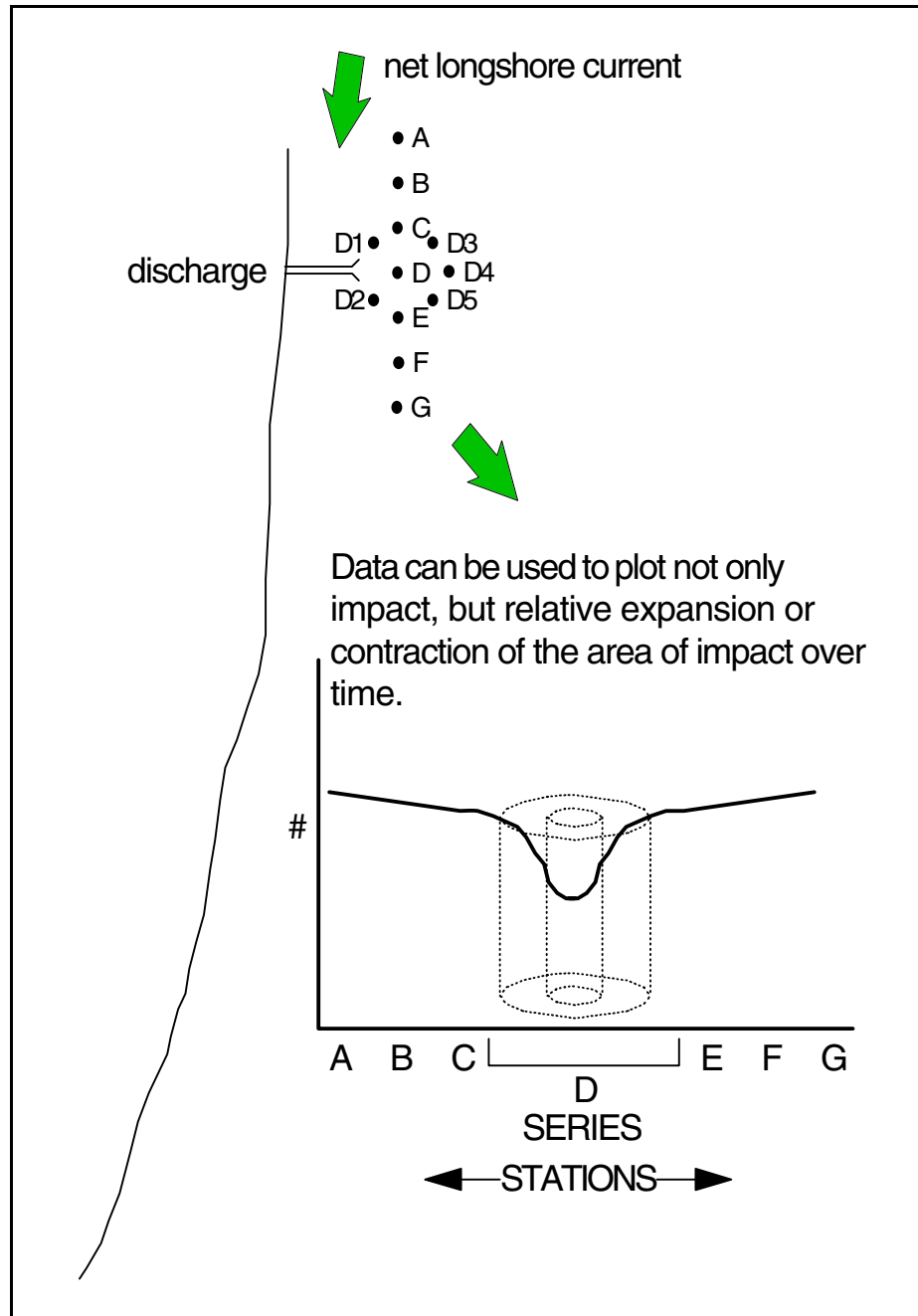
acceptable variation for the region. Thus, the stations (or sites) are considered to all be of comparable habitat characteristics, and because of the spatial arrangement, sites A, E, and I are selected as references.

Reference Condition: The reference condition may be derived from the interquartile range of scores of the values of the biotic condition measured at the reference sites. Table 13-13 presents the range of those values for the summer parameters measured at each of the three reference sites and the mean range of those scores. The range was selected over mean or median values to accommodate the variability of the biological data. This mean range is the reference condition or minimally impacted (by human activities, e.g. sewage discharge, all other factors being considered equal) condition for

Table 13-13. Establishment of reference condition using the mean of the interquartile range of scores for three reference sites.

Station	Individuals	Taxa	Simpson's Index	Shannon-Wiener	Richness
A	427-3049	46-71	0.075-0.161	2.597-3.137	7.3-9.1
E	281-474	40-49	0.076-0.224	2.262-3.889	6.6-8.0
I	136-841	27-42	0.129-0.260	1.993-2.524	5.3-6.0
Mean Score	281-1455	38-54	0.093-0.215	2.284-3.183	6.4-7.7

this Maryland-Delaware reach of the Mid-Atlantic Bight.

Biocriteria: The elements of a biocriterion are: (1) historical information about the area; (2) present reference condition information; (3) empirical modeling of data if needed; and (4) an assessment of this preceding information by a locally familiar panel of specialists.

(1) There is insufficient local historical information or data to establish a trend against which the reference condition data can be compared.

(2) The present reference condition data is presented above.

(3) The indexes used to compile the raw data constitute the only modeling element since this is a site specific assessment.

(4) The authors of this manual are surrogates for a panel of local specialists which would likely consist of USEPA, US Fish and Wildlife Service, NOAA, and State biologists and water resource managers.

Consequently, the reference site data and index scores presented here essentially comprise by default, the candidate biocriteria for the purposes of this study. However, it is important to note that the other elements of development of a biocriterion should not be casually dismissed. While the reference condition is essential, with a large available historical database these present values might well be adjusted either up or down to accommodate the historical trend for the area.

Assessment Comparing Biocriteria to the Test Sites: The test sites at "C" (Bethany Beach, DE outfall) and at "G" (Ocean City, MD outfall) are then compared to the biocriteria as illustrated in Table 13-14.

Neither outfall site completely meets the range of criteria derived from the reference condition for any of the metrics applied, although the Bethany Beach outfall approximates the criterion for number of taxa present. It should be noted that the outlier at reference station "A" (perhaps caused by Delaware Bay enrichment) raises this criterion range at the expense of the Bethany Beach outfall. The Ocean City outfall nearly fits the diversity index criterion. However the outfall far exceeds the number of individuals category by more than three times the criterion. This reflects several instances when the benthic grab was overwhelmed by polychaete worms, a condition usually indicative of sewage pollution.

Table 13-14. Comparison of the reference condition derived biocriteria to the interquartile range of scores at the Bethany Beach and Ocean City outfalls.

Variable	Biocriteria	Bethany Beach Outfall	Ocean City Outfall
No. Indiv.	281-1455	260-1988	49-6,492
No. Taxa	=/>38-54	28-43	13-49
Simps. Dom.	=/<0.093-0.215	0.171-0.642	0.179-0.643
Shan.-Wien. DI	=/>2.284-3.183	0.970-2.648	1.855-2.883
Richness	=/>6.4-7.7	4.6-5.8	3.1-5.7

This instance of near exceedance of one of the criteria in each case illustrates the importance of using several biological metrics to establish a reference condition which best represents a diverse and healthy community, and which contributes to more robust and sensitive biocriteria.

Conclusion: For a formal criteria development program, more data are required, but the indexes applied appear to ably translate the data into workable criteria. Ironically, the number of individuals and number of taxa metrics individually do not reflect apparent conditions as well as the indexes which combine these primary variables.

Recommendations: The stations should continue to be monitored by USEPA Region III biologists and the data set further developed. The long term areal impact of the discharges should be better assessed using the additional near-discharge stations described above. Changes in sieve size and use of grab fractions, if justified, will help reduce the cost of the monitoring.

Eventually, as a further cost reduction measure, it may be possible to monitor just stations "A", "C", "E", "G", and "I". However, periodically the outfall stations should be intensively monitored to determine if the zones of impact are expanding or contracting. The combined information of criteria comparisons and impact zone measurements should provide valuable information for NPDES permit evaluations at Bethany Beach and Ocean City.

This technique and evaluation approach may prove particularly helpful as Eastern Seaboard development continues to increase and more coastal communities seek ocean discharge permits for their municipal effluents.

Primary Contact: George Gibson, Jr. (4304), USEPA, Office of Water, Office of Science and Technology
1200 Pennsylvania Avenue, NW
Washington, DC 20460
410-305-2618
gibson.george@epa.gov

13.7 Environmental Quality of Estuaries of the Carolinian Province: 1995

13.7.1 Background/Objectives

A study was conducted to assess the environmental condition of estuaries in the EMAP Carolinian Province (Cape Henry, VA - St. Lucie Inlet, FL; Figure 13-14). The objectives of this study are being addressed using a probability-based sampling design, under which a large regionally extensive population of randomly selected sites is sampled from year to year, following earlier EMAP-E designs (Strobel et al. 1994, Summers et al. 1993). This design makes it possible to produce unbiased estimates of the percent area of degraded vs. undegraded estuaries, based on a series of indicators of environmental quality. Overall, the objectives of the program are to:

▸ Assess the condition of estuarine resources of the Carolinian Province based on a variety of synoptically measured indicators of environmental quality;

▸ Establish a baseline for evaluating how the condition of these resources are changing with time;

▸ Develop and validate improved methods for use in future coastal monitoring and assessment efforts.

A total of 87 randomly located stations were sampled from July 5 - September 14, 1995 in accordance with the probabilistic sampling design. Wherever possible, synoptic measures were made of: 1) general physical habitat condition, 2) pollution exposure, 3) biotic conditions, and 4) aesthetic quality. Percentages of degraded vs. undegraded estuarine area were compared to results of a related EMAP survey conducted in 1994 in this same region as part of a multi-year monitoring effort.

13.7.2 Methods

An overall goal of EMAP is to make statistically unbiased estimates of ecological condition with known confidence. To approach this goal, a probabilistic sampling framework was established among the overall population of estuaries comprising the Carolinian Province. Under this design, each sampling point is a statistically valid probability-based sample. Thus, percentages of estuarine area with values of selected indicators above or below suggested environmental guidelines can be estimated based on the conditions observed at individual sampling points. Statistical confidence intervals around these estimates also can be calculated. Moreover, these estimates can be combined with those for other regions that were sampled in a consistent manner to yield national estimates of estuarine condition. This section describes in brief how stations were selected using the probabilistic design (see also Rathbun 1994). Supplemental sites, selected non-randomly in clean areas and in suspected polluted areas, were included in the survey and are discussed below.

Sampling sites in 1995 consisted of 87 base stations and 21 supplemental stations. Base stations were randomly selected sites that made up the probability-based monitoring design. Four replicate bottom grabs were collected from each station with a 0.04-m^2 young grab sampler. Data collected

Figure 13-14

1995 Carolinian Province sampling stations.
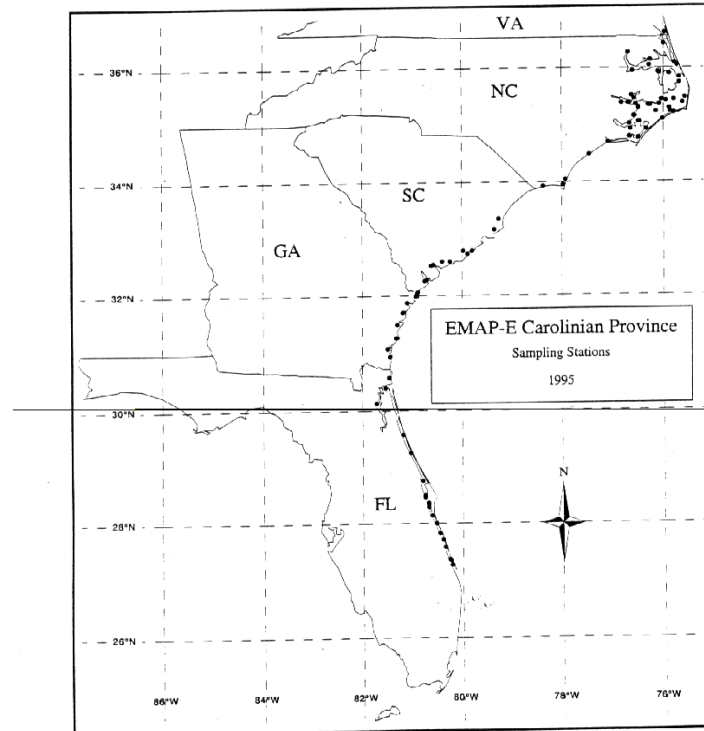
from these sites were used to produce unbiased estimates of estuarine condition throughout the province based on the various synoptically measured indicators of environmental quality. The province-wide distribution of base sites is shown in Figure 13-14. Supplemental stations were selected non-randomly in areas for which there was some prior knowledge of the ambient environmental conditions. These sites, which represented both pristine areas and places with histories of anthropogenic disturbance, were used to test the discriminatory power of various ecological indicators included in the program. Data from supplemental sites were not included in the probabilistic spatial estimates.

As in other EMAP-E provinces (Strobel et al. 1994, Summers et al. 1993), the sampling design for the base sites in the Carolinian Province was stratified based

mainly on the physical dimensions of an estuary. Table 13-15 breaks down the estuarine resources of the Carolinian Province by their size designation. Stratification of the overall sampling area into classes of estuaries with similar attributes was necessary in order to minimize within-class sampling variability. Also, it was not feasible to sample all of the different types of estuaries that exist within a broad geographic region at the same spatial scale. Stratification by physical dimensions of an estuary was adopted because: 1) such attributes usually show minimal change over extended periods; 2) alternative classification variables such as salinity, sediment type, depth, and extent of pollutant loadings would result in the definition of classes for which areal extents could vary widely from year to year; 3) data for physically based classes can be aggregated into geographic units that are meaningful

Table 13-15. Estuarine resources of the Carolinian Province.

	Province	Large Estuaries	Small Estuaries	Large Tidal Rivers
All Years				
Number of Estuaries	200	3	194	3
Area Represented (km^2)	11,622.1	5,581.1	4,907	1,134
In 1995				
Number of Stations	88	16	55[a]	17[b]
Area Represented (km^2)	6,991.8	4,480.0	1,377.8	1,134

[a]Station count includes 6 replicate stations
[b]Station count includes 3 replicate stations

from a regulatory or general-interest perspective; and 4) estuarine boundaries can be delineated more readily and accurately from maps or charts of the physical dimensions of coastal areas than from maps of sediment or water-column characteristics.

Selection of base-site sampling approaches varied on the physical characteristics of the particular estuary being sampled. Base sites in all estuaries were selected using an approach similar to the one used in the EMAP Louisianian Province (Summers et al. 1993). In large estuaries, sites were selected using a sampling grid approach. A triangular lattice was placed initially over the study region and the resulting grid shifted randomly. In large tidal rivers, base sites were selected randomly, using a "spine and rib" approach. Finally, base sites in small estuaries were selected using a random list-frame approach, also similar to the approach used in the EMAP Louisianian Province (Summers et al. 1993). Table 13-16 lists the core environmental indicators sampled at the various sites.
A standard series of environmental parameters was measured at each of the

base stations to provide a consistent set of synoptic data for making province-wide estimates of estuarine condition. These "core" environmental indicators included measures of general habitat conditions, pollutant exposure, biotic integrity, and aesthetic quality (Table 13-16). Habitat indicators describe the physical and chemical conditions of sample sites, and provide basic information about the overall environmental setting. Exposure indicators provide measures of the types and amounts of pollutants, or other adverse conditions, that could be harmful to resident biota or human health. Biotic condition indicators provide measures of the status of biological resources in response to the surrounding environmental conditions. Aesthetic indicators provide additional measures of environmental quality from a human perceptual perspective. There is a fair amount of overlap among these various indicator categories. For example, some aesthetic indicators (presence of oil sheens, noxious sediment odors, and highly turbid waters) could also reflect adverse exposure conditions. Another example is dissolved oxygen (DO), listed as an exposure indicator because of the

Table 13-16. Core environmental indicators for the Carolinian Province.

Habitat Indicators

Water depth
Water temperature
Salinity
Density stratification of water column
Dissolved oxygen concentrations
pH
Percent silt-clay content of sediments
Percent TOC in sediments
Sediment acid-volatile sulfides (Yr. 2 only[*])

Exposure Indicators

Low dissolved oxygen conditions
Sediment contaminants
Contaminants in fishes and invertebrates (Yr. 2 only)
Sediment toxicity

Biotic Condition Indicators

Infaunal species composition
Infaunal species richness and diversity
Infaunal abundance
Benthic Infaunal Index
Demersal species composition (invertebrates and fish)
Demersal species richness and diversity
Demersal species abundance
Demersal species lengths
External pathological abnormalities in demersal biota

Aesthetic Indicators

Water clarity
Anthropogenic debris (sea surface and in trawls)
Noxious sediment odors (sulfides, petroleum)
Oil sheens (sea surface and bottom sediments)
[*]Results not shown in this report

potential adverse biological effects of low oxygen concentrations, but which also is clearly a measure of general habitat conditions. These various core environmental parameters included ones used in other EMAP-E provinces (Strobel et al. 1994, Summers et al. 1993) to support regional comparisons and to provide a means for producing combined nationwide estimates of estuarine condition.

In addition to making the standard EMAP-E measurements, an emphasis was placed on developing and validating other complementary methods to aid in evaluating the quality of southeastern estuaries. Such indicators, some still in the development stage, are listed in Table 13-17. They include sediment bioassays with alternative test species, such as the amphipod *Ampelisca verrilli* as an alternative to *A. abdita* in standard 10-day solid-phase toxicity tests; assays with additional sublethal biological endpoints, such as effects on feeding, growth and fertilization success in key estuarine organisms; additional indices of environmental quality for tidal marshes and estuarine fish assemblages; and the incorporation of additional exposure indicators, such as porewater ammonia and hydrogen sulfide concentrations, to help in the

Table 13-17. Exposure indicators under development in the Carolinian Province.

10-day acute-toxicity sediment bioassay with alternative amphipod species, *Ampelisca verrilli*

1-week sublethal bioassay for testing effects of sediment exposure on growth of juvenile clams *Mercenaria mercenaria*

96-hour sublethal bioassay for testing effects of sediment exposure on feeding rates of *Ampelisca verrilli*

1-hour sublethal bioassay using gametes of oysters *Crassostrea virginica* and clams *Mercenaria mercenaria* for testing effects of sediment exposure on fertilization success

Sediment porewater ammonia and hydrogen sulfide concentrations

---

interpretation of sediment toxicity results.

13.7.3 Benthic Infaunal Index

The modified IBI approach of Weisberg et al. (1997) was used to develop a benthic index for southeastern estuaries. The goal was to develop an index that possessed the following features: (1) suitable for use throughout the region, (2) applicable to a broad range of habitats, (3) easy to understand and interpret, and (4) effective in discriminating between undisturbed and disturbed conditions associated with human influences.

Results of the 1994 survey (Hyland et al. 1996) indicated that several natural abiotic factors (salinity, latitude, silt-clay, and TOC) had strong influences on infaunal variables. In the IBI approach, an attempt is made to account for such variations by defining habitat-specific reference conditions at sites free of anthropogenic stress and then comparing conditions in samples with the expected reference conditions for similar habitat types. The basic steps used to develop the index involved: (1) defining major habitat types based on classification analysis of benthic species composition and evaluation of the physical characteristics of the resulting site groups; (2) selecting a development data set representative of degraded and undegraded sites in each habitat (3) comparing various benthic attributes between reference sites and degraded sites for each of the major habitat types; (4) selecting the benthic attributes that best discriminated between reference and degraded sites for inclusion in the index; (5) establishing scoring criteria (thresholds) for the selected attributes based on the distribution of values at reference sites; (6) constructing a combined index value for any given sample by assigning an individual score for each attribute, based on the scoring criteria, and then averaging the individual scores; and (7) validating the index with an independent data set.

Data from undegraded sites sampled in 1993 and 1994 were first analyzed using classification (cluster) analysis of benthic species composition and evaluation of the physical factors associated with the resulting station clusters to define major habitat types. Several types of cluster analyses were performed. The one that produced the clearest results was a normal (Q-mode) analysis run on log10-transformed data with flexible sorting as the clustering method and Bray-Curtis similarity as a resemblance measure (see Boesch 1977).

Differences in abiotic factors (salinity, latitude, % silt-clay, TOC) among the resulting station clusters were examined by ANOVA and pair-wise multiple comparison tests (Duncan's test and Tukey's HSD) to help delineate the major habitat types. Four site groups resulted: oligohaline-mesohaline stations ($\leq 8\%$) from all latitudes, polyhaline-euhaline stations ($>18\%$) from northern latitudes ($>34.5°$ N), polyhaline-euhaline stations from middle latitudes (30-$34.5°$ N) and polyhaline-euhaline stations from southern latitudes ($<30°$ N). Seventy-five stations sampled during the 1994 survey were selected for the development data set. These stations provided data from both degraded and undegraded sites in each of the four habitats. Classification of stations into degraded and undegraded categories was based on the combination of chemical and toxicological criteria, mainly DO, and toxicity of sediment bioassays. Marginal sites (minor evidence of stress with toxicity in only one assay and no accompanying adverse contaminant or DO conditions) were not included in the development data set.

Forty different infaunal attributes were tested with the 1994 development data set to determine those that best discriminated between undegraded and degraded sites within each habitat. This initial list of attributes included various measures of diversity, abundance, dominance, and presence of indicator species (e.g., pollution-sensitive vs. pollution-tolerant species, surface vs. subsurface feeders). A subset of six candidate metrics was identified for possible inclusion in the index. Key criteria considered in the selection were whether differences were in the right direction and statistically significant (based on results of Student t-tests, Mann-Whitney U-tests, and Komogorov-Smirnov two-sample tests;

at $\alpha = 0.1$). These six metrics were: mean number of taxa, mean abundance (all taxa), mean H' diversity, 100 - % abundance of the two most numerically dominant species, and two different measures of % abundance of pollution-sensitive taxa.

Scoring criteria for each of these metrics were developed based on the distribution of values at undegraded sites: score of 1, if value of metric for sample being evaluated was in the lower 10th percentile of corresponding reference-site values; score of 3, if value of metric for sample was in the lower 10th-50th percentile of reference-site values; or score of 5, if value of metric for sample was in the upper 50th percentile of reference-site values. Scoring criteria were determined separately for each metric and habitat type. A combined index value was then computed for a sample by assigning a score for each component metric (based on the individual scoring criteria for the corresponding habitat type) and then averaging the individual scores. A combined score < 3 suggested the presence of a degraded benthic assemblage (some apparent level of stress to very unhealthy) given that its condition, based on the averaged metrics, deviated from conditions typical of the "best" (upper 50th percentile) reference sites.

Forty different combinations of the six candidate benthic metrics were further evaluated to determine which represented the best combined index. The metric combination that produced the highest percentage of correct classifications; i.e., agreement with predictions of sediment bioeffects based on the chemistry and toxicity data, was then selected to represent the final index. The resulting final index was the average score of four metrics: (1) mean abundance, (2) mean number of taxa, (3)

100 - % abundance of the top two numerical dominants, and (4) % abundance of pollution-sensitive taxa (i.e., percent of total faunal abundance represented by Ampeliscidae + Haustoriidae + Hesionidae + Tellinidae + Lucinidae + Cirratulidae + *Cyathura polita* + *C. burbanki.* The final combined index correctly classified 93% of the stations province-wide in the development data set and 75% of the stations in the independent validation data set.

13.7.4   Results

The multimetric index of biotic integrity index — consisting of measures of abundance, number of species, dominance, and relative abundance of pollution-sensitive taxa — produced a high percentage of correct station classifications; i.e., agreement with predictions of sediment bioeffects based on chemistry and toxicity data, in comparison to other metric combinations that were tested. The index correctly classified stations province-wide 93% of the time in the 1994 development data set and 75% of the time in the independent 1993/1995 validation data set.

Figure 13-15 illustrates that stations with index values below 3 (suggestive of some apparent stress to highly degraded conditions) usually coincided with sites considered to be degraded based on a combination of chemistry and toxicity data, and that stations with scores of 3 or higher usually coincided with undegraded sites. Agreement is the highest at the two ends of the scale. Thus, the evaluation of sediment quality based on the benthic index appears to agree reasonably well with predictions of sediment bioeffects based on the combined exposure data. Additional comparisons revealed that the benthic index detected a higher

percentage of samples where bioeffects were expected (based on sediment quality guideline exceedances) than did any of the four individual sediment bioassays (Fig. 13-16a) or individual infaunal attributes (Fig. 13-16b). Benthic index values for base stations sampled in 1995 covered the full scale from 1 to 5. Values < 1.5 (clearest evidence of a degraded benthos) occurred at 14 of the 86 base sites, which represented 21% of the province area (Fig.13-17). Transitional values of 2 to 2.5 (suggestive of some possible stress) occurred at an additional 14 sites, representing another 15% of the province. Values ≥3 (suggestive of an undegraded benthos) occurred at the remaining 58 base sites, representing 64% of the area of the province.

By estuarine class, the estimated percentage of area with degraded benthic assemblages was the highest for large tidal rivers and the lowest for large estuaries (Fig. 13-18). By subregion, this percentage was the highest in Florida estuaries and the lowest in South Carolina/Georgia estuaries.

Extracted or summarized from the EMAP Carolinian Province Report, Annual Statistical Summary for the 1995 EMAP - Estuaries Demonstration Project in the Carolinian Province (Hyland et al. 1998).

Primary Contact: Jeffrey L. Hyland, Carolinian Province Office, NOAA/National Ocean Service
217 Fort Johnson Rd. (P.O. Box 12559), Charleston, SC 29422-2559
843-762-5415
jhyland@rdc.noaa.gov

Figure 13-15

Frequency distribution of index scores for undegraded vs. degraded stations in 1993/1995 "development" data set.
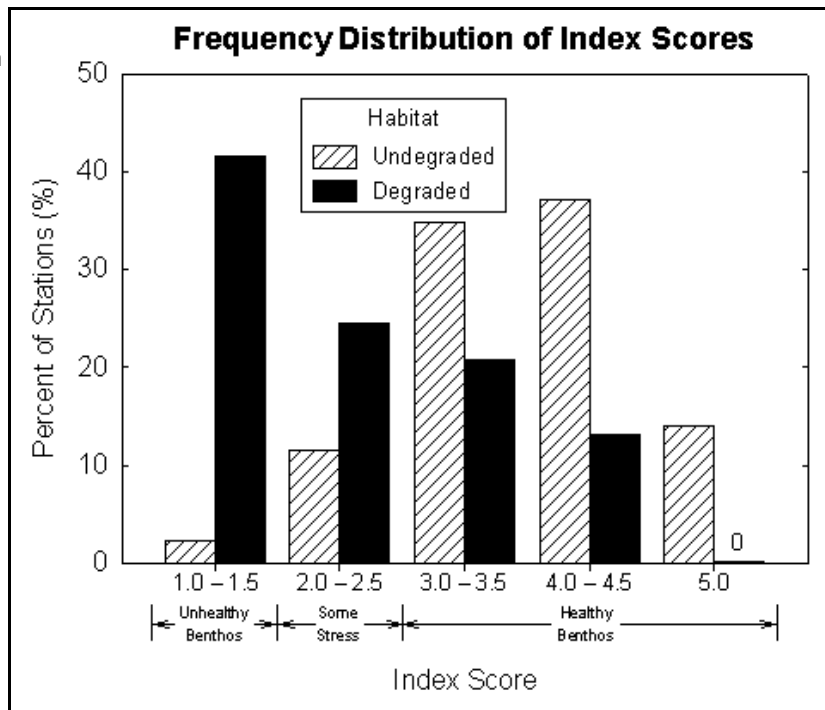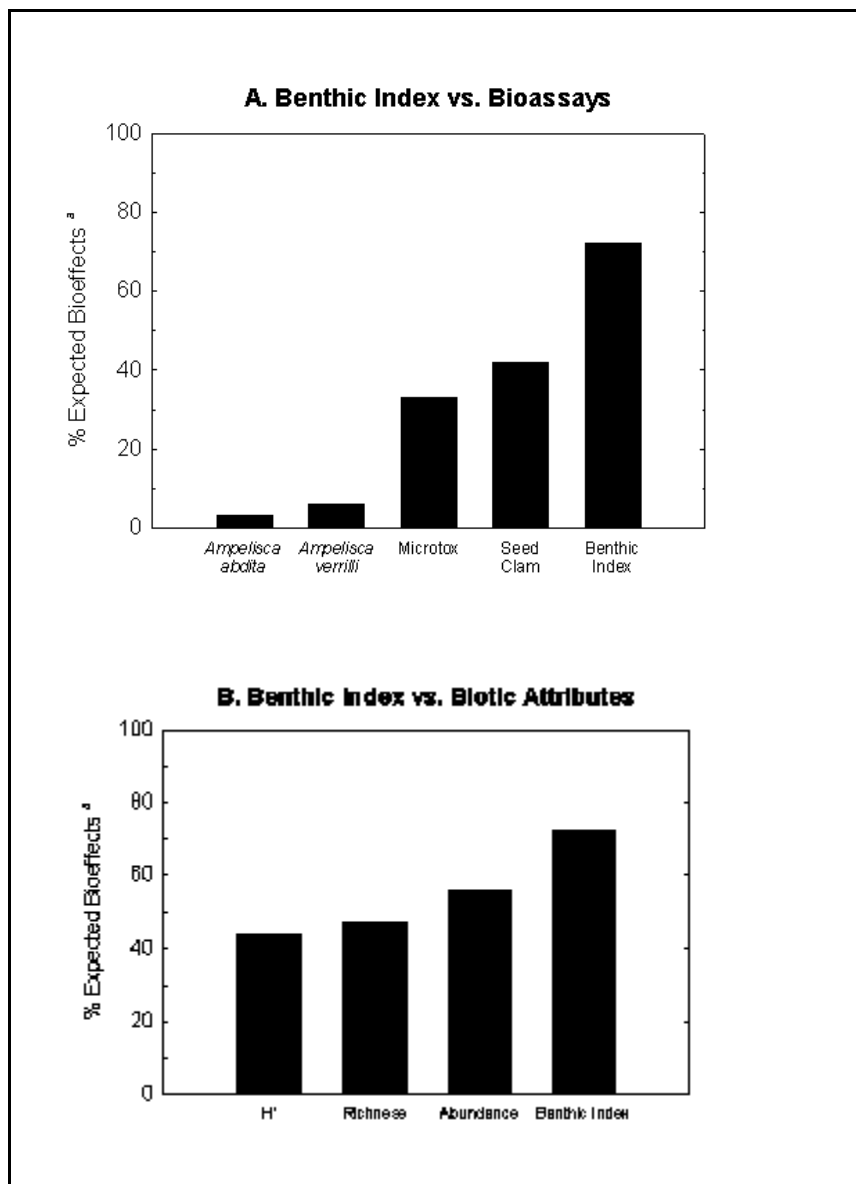
Figure 13-16

Comparison of the percent of expected bioeffects detected with the benthic index vs. (A) four sediment bioassays and (B) three individual infaunal attributes. [a]Percent expected bioeffects - # stations (1995 core & supplemental) where an effect was detected / # stations with ≥ 1 ER-M/PEL or ≥ 3 ER-L/TEL exceedance.

Figure 13-17

Percent area
(and 95% C.I.) of
CP estuaries with
high (≥ 3),
intermediate (>
1.5 to < 3), and
low (≤ 1.5)
benthic index
values.
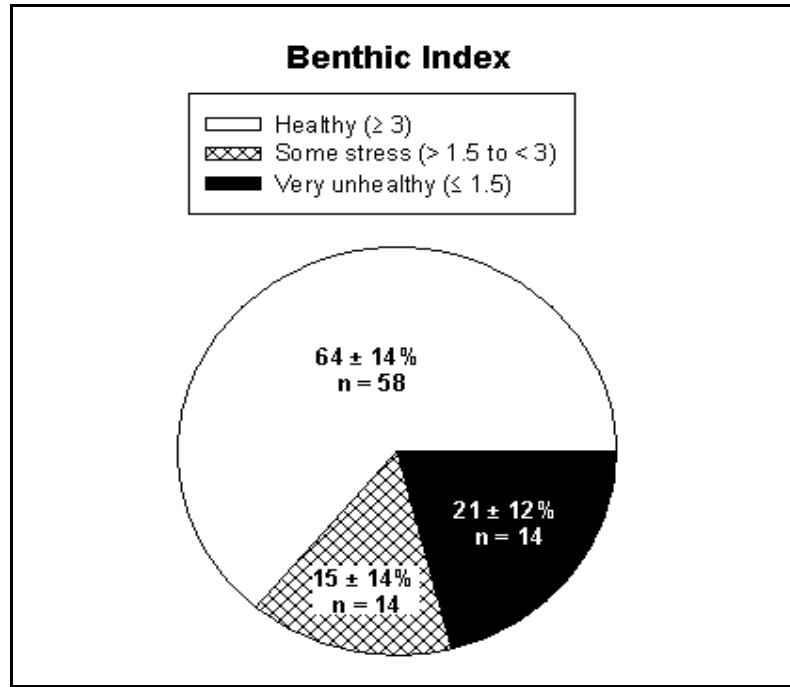
Figure 13-18

Comparison of
benthic index
values by
estuarine class
and subregion.
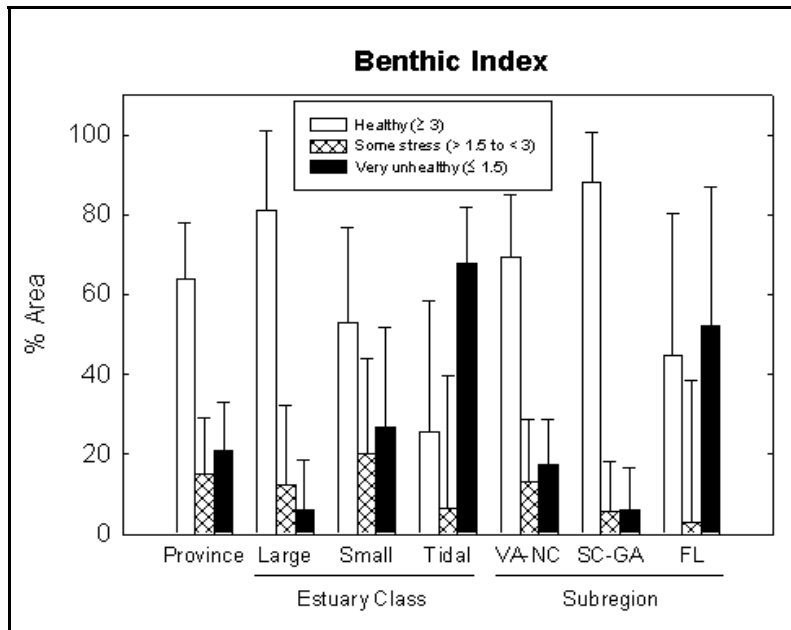
13.8 Assessment of the Ecological Condition of the Delaware and Maryland Coastal Bays

13.8.1 Background

The coastal bays formed by the barrier islands of Maryland and Delaware are important ecological and economic resources whose physical characteristics and location make them particularly vulnerable to the effects of pollutants. A first step in developing management strategies for these systems is to characterize their present condition and how it has changed over time. This project was undertaken as a collaborative effort of the Coastal Bays Joint Assessment (CBJA), a group of state and federal agencies, to assess the ecological condition of this system and fill a data void identified in previous characterization studies.

Two hundred sites were sampled in the summer of 1993 using a probability-based sampling design that was stratified to allow assessments of the coastal bays as a whole, each of four major subsystems within coastal bays (Rehoboth Bay, Indian River Bay, Assawoman Bay, and Chincoteague Bay) and four target areas of special interest to resource managers (upper Indian River, St. Martin River, Trappe Creek, and artificial lagoons). Measures of biological response, sediment contaminants, and eutrophication were collected at each site using the same sampling methodologies and quality assurance/quality control procedures used by EMAP. The consistency of the sampling design and methodologies between this study and EMAP allows unbiased comparison of conditions in the coastal bays with that in other major estuarine systems in USEPA Region III that are sampled by EMAP. As an additional part of the study, trends in fish communities structure were assessed by collecting monthly beach seine and trawl measurements during the summer at about 70 sites where historic measurements of fish communities have been made.

13.8.2 Methods

Sampling sites were selected using a stratified random sampling design in which the coastal bays were stratified into several subsystems for which independent estimates of condition were desired:

▸ Upper Indian River;

▸ Trappe Creek/Newport Bay;

▸ St. Martin River;

▸ Artificial lagoons throughout the coastal bays;

▸ All remaining areas within Maryland's coastal bays; and

▸ All remaining areas within Delaware's coastal bays.

The upper Indian River, Trappe Creek, and St. Martin River were defined as sampling strata because resource managers expressed particular concern about these areas. Water quality data suggest that each of these tidal creeks is subject to excessive nutrient enrichment, algal blooms, and low concentrations of DO. These creeks are also believed to transmit large nutrient loads (from agricultural runoff) downstream contributing to eutrophication throughout the coastal bays (Boynton et al. 1993).

Artificial lagoons were defined as a stratum because of their high potential for impact based on their physical

characteristics and their proximity to a variety of contaminant sources (Brenum 1976). These dredged canal systems can form the aquatic equivalent of streets in development parcels; they already encompass 105 linear miles and almost 4% of the surface area of Delaware's inland bays. In general, these systems are constructed as dead-end systems with little or no freshwater inflows for flushing. They are often dredged to a depth greater than the surrounding waters, leaving a ledge that further inhibits exchange with nearby waters and leads to stagnant water in the canals. The placement of these systems in relatively high density residential areas increases the potential contaminant input. Much of the modified land-use in dredged canal systems extends to the edge of the bulkheaded waters, providing a ready source of unfiltered runoff of lawn-care and pesticides. In many cases, the bulkhead and dock systems in these canal systems are built from treated lumber containing chromium, copper, and arsenic, providing another source of contaminants.

Four replicate bottom grabs were collected from each station with a 0.04-m^2 Young grab sampler. Of the two hundred sites sampled, 25 were in each of the first four sampling strata and 50 were in each of the last two. Sites were selected by simple random sampling in all strata except artificial lagoons. The randomly selected sites were chosen by enhancing the base EMAP grid (Overton et al. 1990). A different level of enhancement was applied to each stratum to obtain the required number of samples. Sites in the artificial lagoons were selected by developing a list frame (of all existing lagoons), randomly selecting 25 lagoons from that list, and then randomly selecting a site within each selected lagoon.

All sampling was conducted between July 12 and September 30, 1993. Sampling was limited to a single index period because available resources were insufficient to sample in all seasons. Late summer is the time during which environmental stress on estuarine systems in the mid-Atlantic region is expected to be greatest owing to high temperatures and low dilution flows (Holland 1990). The sampling period coincided with the period during which EMAP sampled estuaries of the mid-Atlantic region; therefore, data collected in the coastal bays annually for EMAP can be incorporated into estimates of ecological condition generated from Coastal Bays Joint Assessment (CBJA) data. That data can then contribute to continuing development and evaluation of EMAP indicators.

Measurements of physical characteristics provide basic information about the natural environment. Knowledge of the physical context in which biological and chemical data are collected is important for interpreting results accurately because physical characteristics of the environment determine the distribution and species composition of estuarine communities, particularly assemblages of benthic macroinvertebrates. Salinity, sediment type, and depth are all important influences on benthic assemblages (Snelgrove and Butman 1994, Holland et al. 1989). Sediment grain size also affects the accumulation of contaminants in sediments. Fine-grained sediments generally are more susceptible to contamination than sands because of the greater surface area of fine particles (Rhoads 1974, Plumb 1981).

Depth, silt-clay content of the sediment, bottom salinity, temperature, and pH were measured to describe the physical conditions at sites in the coastal bays.

Sediment type was defined according to silt-clay content (fraction less than 63-μ); classifications were the same as those used for EMAP. Biologically meaningful salinity classes were defined according to a modified Venice System (Symposium on the Classification of Brackish Waters 1958).

Healthy aquatic ecosystems require clear water, acceptable concentrations of dissolved oxygen, limited concentrations of phytoplankton, and appropriate concentrations of nutrients. Clear water is a critical requirement for submerged aquatic vegetation (SAV), which provides habitat for many other aquatic organisms (Dennison et al. 1993). As large concentrations of suspended sediment or algal blooms reduce water clarity, the amount of sunlight reaching SAV is diminished and the plants fail to thrive; consequently, critical habitat for crabs, fish, and other aquatic organisms is lost (Dennison et al. 1993). Nutrient enrichment causes excessive algal growth in the water column and on the surfaces of plants. As bacteria metabolize the excess algae, they deplete dissolved oxygen in the water column and sediments causing hypoxia and, in extreme cases, anoxia.

Water quality in the coastal bays of Delaware and Maryland was evaluated using classes of indicators: measures of algal productivity, dissolved oxygen (DO), water clarity, and nutrients. Measures of algal biomass included the concentrations of chlorophyll in the water column and sediment, and phaeophytin. Secchi depth, total suspended solids (TSS), and turbidity were measured to assess water clarity. Nutrient measures included dissolved inorganic nitrogen (DIN; nitrite, nitrate, and ammonium), dissolved inorganic phosphorus (DIP), total dissolved nitrogen (TDN), total dissolved

phosphorus (TDP), and particulate nitrogen and phosphorus. Table 13-18 lists the core environmental parameters sampled at the various sites.

Estimating the percent of eutrophied area in the coastal bays requires identifying threshold levels for selected indicators that define eutrophication. While no such levels have been established for the coastal bays, the Chesapeake Bay Program has established thresholds for five water quality parameters to define critical habitat requirements for supporting SAV in a polyhaline environment (Dennison et al. 1993); these thresholds were used for our assessment (Table 13-19). All but one of the SAV restoration goal attributes were measured directly. The light attenuation coefficient was calculated from Secchi depth measurements.

Threshold values of sediment contaminants developed by Long and Morgan (1990) and updated by Long et al. (1995) were used to interpret concentrations of sediment contaminants measured in the coastal bays. Two values were identified for each contaminant: an effects range-low (ER-L) value corresponding to contaminant concentrations above which biological effects begin to appear, and an effects-range median (ER-M) concentration, above which biological effects are probable. Only a subset of the contaminant samples collected for the CBJA were processed because of cost constraints; consequently, comparisons were limited to the artificial lagoons and the coastal bays as a whole.

Sediment samples for analysis of benthic macroinvertebrates, silt-clay content, benthic chlorophyll, and chemical contaminants were collected using a 0.044-m^2, stainless steel, Young-modified VanVeen grab. Four measures of

Table 13-18. Environmental parameters for the Maryland/Delaware Coastal Bays.

Physical Parameters
Depth
% Silt/Clay content
Salinity
Temperature
pH
Water Quality Parameters
Chlorophyll *a*
Phaeophytin
Benthic chlorophyll
DO (Dissolved Oxygen)
NO_2 (Nitrite)
NO_3 (Nitrate)
Ammonium
TDN (Total Dissolved Nitrogen)
Orthophosphate
TDP (Total Dissolved Phosphorus)
TPN (Total Particulate Nitrogen)
TPP (Total Particulate Phosphorus)
TPC (Total Particulate Carbon)
Secchi Depth
TSS (Total Suspended Solids)
Turbidity
Benthic Parameters
Abundance
Biomass
Number of Species
Shannon-Wiener Index
EMAP Index

Table 13-19. Chesapeake Bay submerged aquatic vegetation habitat requirements for a polyhaline environment (Dennison et al. 1993).

Parameter	Critical Value
Light attenuation coefficient (k_d; m^{-1})	1.5
Total suspended solids (mgL^{-1})	15
Chlorophyll *a* (µg/l)	15
Dissolved inorganic nitrogen (µM)	10
Dissolved inorganic phosphorus (µM)	0.67

biological response were used to evaluate the condition of benthic assemblages in the coastal bays of Delaware and Maryland: abundance, biomass, diversity, and the EMAP benthic index. Abundance and biomass are measures of total biological activity at a location. The diversity of benthic organisms supported by the habitat at a location often is considered a measure of the relative "health" of the environment. Diversity was evaluated using the number of species; i.e., species richness, at a location and the Shannon-Wiener diversity index, which incorporates both species richness and evenness components. The EMAP benthic index integrates measures of species diversity, composition, biomass, and abundance into a single value that distinguishes between sites of good or poor ecological condition (Schimmel et al. 1994). A value of 0 or less denotes a degraded site at which the structure of the benthic community is poor, and the number of species, abundance of selected indicator species, and mean biomass are small.

13.8.3   Results/Conclusions

Major portions of the coastal bays have degraded environmental quality. EMAP's benthic index measured 28% of the area in the coastal bays had degraded benthic communities. At least one sediment contaminant exceeding the Long et al. (1995) ER-L concentration (threshold of initial biological concern) were found in 68% of the area in the coastal bays. More than 75% of the area in the coastal bays failed the Chesapeake Bay Program's Submerged Aquatic Vegetation (SAV) restoration goals, which are a combination of measures that integrate nutrient, chlorophyll, and water clarity parameters.

The tributaries to the coastal bays are in poorer condition than the mainstems of the major subsystems. Previous studies have suggested that the major tributaries to the system: upper Indian River, St. Martin River, and Trappe Creek are in poorer condition than the mainstem water bodies. This study confirmed that finding. The percentage of area containing degraded benthos was generally two to three times greater in the tributaries compared to the rest of the coastal bays. The percent of area with DO less than the state standard of 5-ppm was three to seven times greater in the tributaries. More than 70% of the area in upper Indian River and St. Martin River and in the artificial lagoons had chlorophyll *a* concentrations exceeding the SAV restoration goals.

Among these systems, Trappe Creek contained the sites in the worst condition. Two sites in the upper portion of Trappe Creek had concentrations of chlorophyll *a* exceeding 350 μgL^{-1}; algal blooms were evident at each site. In addition, daytime DO levels exceeding 14-ppm were measured at both sites. Although, supersaturated DO often occurs in hypereutrophic waterbodies on warm, sunny days. However, it appears that degraded conditions in the Trappe Creek system are spatially limited to Trappe Creek and have not spread to Newport Bay. Undoubtedly, this results from the low freshwater flow from this tributary compared to the other tributaries.

Moreover, the coastal bays are in as poor or worse condition than either the Chesapeake or Delaware Bays with respect to sediment contaminant levels, water quality, and benthic macroinvertebrate community

condition. Based on comparison to EMAP data collected between 1990-1993, the coastal bays were found to have 68% chemical contamination in the sediments, a higher prevalence than either Chesapeake Bay or Delaware Bay. The total area in the coastal bays that had at least one sediment contaminant exceeding the Long et al. (1995) ER-L concentration was 50% higher than the spatial extent EMAP estimated for Chesapeake Bay using identical methods, and 40% higher, though not statistically distinguishable, from what EMAP estimated for Delaware Bay.

Twenty-eight percent of the area in the coastal bays had degraded benthic communities as measured by EMAP's benthic index. This was significantly greater than the 16% EMAP estimated for Delaware Bay using the same methods and same index, and statistically indistinguishable from the 26% estimated for the Chesapeake Bay.

Nutrients were not measured by EMAP and statistically unbiased estimates of average concentrations are unavailable for either Chesapeake or Delaware Bays. The Chesapeake Bay Program though, recently estimated that about 75% of the area in Chesapeake Bay meets SAV Restoration Goals. This is more than three times the percent of area meeting SAV Restoration Goals in the coastal bays. Even when the turbidity and TSS components of the SAV Restoration Goals (which are naturally high in shallow systems), are ignored, almost half of the area in the coastal bays still fails the SAV Restoration Goal estimates for nutrients and chlorophyll.

The fish community structure in Maryland's coastal bays has remained relatively unchanged during the past twenty years while that of similar systems in Delaware have changed substantially. Fish communities of the Maryland coastal bays are dominated by Atlantic silversides, bay anchovy, Atlantic menhaden, and spot. This community structure is similar to that of the Delaware coastal bays 35 years ago. The fish fauna in Delaware's coastal bays has shifted toward species of the Family Cyprinodontidae (e.g., killifish and sheepshead minnow) which are more tolerant to low oxygen stress, and extremes of salinity and temperature.

Primary Contact: Dr. Frederik W. Kutz
U.S. EPA, Region III
Environmental Science Center
701 Mapes Rd.
 Fort Meade, MD 20755
410-305-2742
Kutz.Rick@epa.gov

Glossary

A posteriori **classification** - a classification based on the results of experimentation.

A priori **classification** - a classification made prior to experimentation.

Aquatic community - an association of interacting populations of aquatic organisms in a given waterbody or habitat.

Aquatic life uses - a subset of designated uses for high quality waters. As such, they are in need of special protection so that characteristics of their resident biotic communities are identified and protected.

Assemblage - an association of interacting populations of organisms in a given waterbody (e.g., fish assemblage or benthic macroinvertebrate assemblage).

Attribute - physical and biological characteristics of habitats which can be measured or described.

Benthic macroinvertebrates - see benthos.

Benthos - animals without backbones, living in or on the sediments, of a size large enough to be seen by the unaided eye, and which can be retained by a U.S. Standard No. 30 sieve (28 openings/in, 0.595-mm openings). Also referred to as benthic macroinvertebrates, infauna, or macrobenthos.

Bioaccumulation - a process by which chemicals are taken up by aquatic organisms directly from water as well as through exposure via other routes, such as consumption of food and sediment containing the chemicals.

Bioconcentration - a process by which there is a net accumulation of a chemical directly from water into aquatic organisms resulting from simultaneous uptake (e.g., via gill or epithelial tissue) and elimination.

Biological assemblage - a group of phylogenetically (e.g., fish) or ecologically (e.g., benthic macroinvertebrates) related organisms that are part of an aquatic community.

Biological assessment or **Bioassessment** - an evaluation of the condition of a waterbody using biological surveys and other direct measures of the resident biota of the surface waters, in conjunction with biological criteria.

Biological criteria or **Biocriteria** - guidelines or benchmarks adopted by States to evaluate the relative biological integrity of surface waters. Biocriteria are narrative expressions or numerical values that describe biological integrity of aquatic communities inhabiting waters of a given classification or designated aquatic life use.

Biological indicators - plant or animal species or communities with a narrow range of environmental tolerances that may be selected for monitoring because their absence or presence and relative abundances serve as barometers of environmental conditions.

Biological integrity - the condition of the aquatic community inhabiting unimpaired waterbodies of a specified habitat as measured by community structure and function.

Biological monitoring or **Biomonitoring** - multiple, routine biological surveys over time using consistent sampling and analysis methods for detection of changes in biological condition.

Biological survey or **Biosurvey** - collecting, processing and analyzing representative portions of an estuarine or marine community to determine its structure and function.

Biomagnification - the result of the processes of bioconcentration and bioaccumulation by which tissue concentrations of bioaccumulated chemicals increase as the chemical passes up through two or more trophic levels in the food chain.

Biota - plants, animals and other living resources.

Brackish - water with salt content ranging between that of sea water and fresh water; commonly used to refer to oligohaline waters.

Coastal waters - marine waters adjacent to and receiving estuarine discharges and extending seaward over the continental shelf and/or the edge of the U.S. territorial sea.

Community - any group of organisms belonging to a number of different species that co-occur in the same habitat or area; an association of interacting assemblages in a given waterbody.

Demersal - living on or near the bottom of a body of water (e.g., mid-water and bottom-dwelling fish and shellfish, as opposed to surface fish).

Designated uses - descriptions of the optimal use of each waterbody as defined by States including natural fisheries, recreation, transportation, or mixed uses.

Discriminant analysis - a type of multivariate analysis used to distinguish between two groups.

Ecological integrity - the condition of an unimpaired ecosystem as measured by combined chemical, physical (including habitat), and biological attributes.

Ecoregion - geographic regions of ecological similarity defined by similar climate, landform, soil, natural vegetation, hydrology or other ecologically relevant variables.

Effects Range-Low - concentration of a chemical in sediment below which toxic effects were rarely observed among sensitive species (10th percentile of all toxic effects).

Effects Range-Median - concentration of a chemical in sediment above which toxic effects are frequently observed among sensitive species (50th percentile of all toxic effects).

Epibenthos - those animals (usually excluding fishes) living on the top of the sediment surface.

Epifauna - benthic animals living on the sediment or on and among rocks and other structures.

Estuarine or coastal marine classes - classes that reflect basic biological communities and that are based on physical parameters such as salinity, depth, sediment grain size, dissolved oxygen and basin geomorphology.

Estuarine waters - semi-enclosed body of water which has a free connection with the open sea and within which seawater is measurably diluted with fresh water derived from land drainage.

Facultative - capable of adaptive response to varying environments.

Habitat - a place where the physical and biological elements of ecosystems provide an environment and elements of the food, cover and space resources needed for plant and animal survival.

Halocline - a vertical gradient in salinity.

Holoplankton - an aggregate of passively floating, drifting or somewhat motile organisms throughout their entire life cycle.

Hypoxia - the condition of low dissolved oxygen in aquatic systems (typically with a concentration < 2-mgL^{-1} but > 0.5-mgL^{-1}).

IBI or **Index of Biotic Integrity** - a fish community assessment approach that incorporates the zoogeographic, ecosystem, community and population aspects of fisheries biology into a single ecologically-based index of the quality of a water resource.

Impact - a change in the chemical, physical or biological quality or condition of a waterbody caused by external sources.

Impairment - a detrimental effect on the biological integrity of a water body caused by an impact.

Indexes - a usually dimensionless numeric combination of scores derived from biological measures called metrics.

Index period - a sampling period, with selection based on temporal behavior of the indicator(s) and the practical considerations for sampling.

Indicator - characteristics for the environment, both abiotic and biotic, that can provide quantitative information on environmental conditions.

Indicator taxa or **Indicator species** - those organisms whose presence (or absence) at a site is indicative of specific environmental conditions.

Infauna - see benthos.

In situ - measurements taken in the natural environment.

Kurtosis - a measure of the departure of a frequency distribution from a normal distribution, in terms of its relative peakedness or flatness.

Littoral zone - the intertidal zone of the estuarine or seashore; i.e., the shore zone between the highest and lowest tides.

Macrobenthos - see benthos.

Macrofauna - animals of a size large enough to be seen by the unaided eye and which can be retained by a U.S. Standard No. 30 sieve (28 meshes/in, 0.595-mm openings).

Macroinvertebrates - animals without backbones of a size large enough to be seen by the unaided eye and which can be retained by a U.S. Standard No. 30 sieve (28 meshes/in, 0.595-mm openings).

Macrophytes - large aquatic plants that may be rooted, non-rooted, vascular or algiform (such as kelp); including submerged aquatic vegetation, emergent aquatic vegetation, and floating aquatic vegetation.

Meiofauna - small interstitial; i.e., occurring between sediment particles, animals that pass through a 1-mm mesh sieve but are retained by a 0.1-mm mesh.

Meroplankton - organisms that are planktonic only during the larval stage of their life history.

Mesohaline - the estuarine salinity zone with a salinity range of 5-18-ppt.

Metric - a calculated term or enumeration which represents some aspect of biological assemblage structure, function, or other measurable characteristic of the biota that changes in some predictable way in response to impacts to the water body.

Multimetric approach - an analysis technique that uses a combination of several measurable characteristics of the biological assemblage to provide an assessment of the status of water resources.

Multivariate community analysis - statistical methods (e.g., ordination or discriminant analysis) for analyzing physical and biological community data using multiple variables.

NPDES or **National Pollutant Discharge Elimination System** - a permit program under Section 402 of the Clean Water Act that imposes discharge limitations on point sources by basing them on the effluent limitation capabilities of a control technology or on local water quality standards.

Oligohaline - the estuarine salinity zone with a salinity range of 0.5-5-ppt.

Optimal - most favorable point, degree, or amount of something for obtaining a given result; in ecology most natural or minimally disturbed sites.

Pelagic - pertaining to open waters or the organisms which inhabit those waters**.**

Pelagic zone - the area of open water beyond the littoral zone.

Percent fines - in analysis of sediment grain size, the percent of fine (.062-mm) grained fraction of sediment in a sample.

Photic zone - the region in a water body extending from the surface to the depth of light penetration.

Plankton - free-floating or drifting organisms with movements determined by the motion of the water.

Population - an aggregate of interbreeding individuals of a biological species within a specified location.

Pseudoreplication - the repeated measurement of a single experimental unit or sampling unit, with the treatment of the measurements as if they were independent replicates of the sampling unit.

Pycnocline - a zone of marked density gradient.

Reference condition - the chemical, physical or biological quality or condition exhibited at either a single site or an aggregation of sites that represents the least impaired condition of a classification of waters to which the reference condition applies.

Reference sites - minimally impaired locations in similar water bodies and habitat types at which data are collected for comparison with test sites. A separate set of reference sites are defined for each estuarine or coastal marine class.

Replicate - taking more than one sample or performing more than one analysis.

Saprobien system - an ecological classification of a polluted aquatic system that is undergoing self-purification. Classification is based on relative levels of pollution, oxygen concentration and types of indicator microorganisms; i.e., saprophagic microorganisms – feeding on dead or decaying organic matter.

Seiche - a wave that oscillates (for a period of a few minutes to hours) in lakes, bays, lagoons or gulfs as a result of seismic or atmospheric disturbances (e.g., "wind tides").

Simulation models - mathematical models (logical constructs following from first principles and assumptions), statistical models (built from observed relationships between variables), or a combination of the two.

Skewness - the degree of statistical asymmetry (or departure from symmetry) of a population. Positive or negative skewness indicates the presence of a long, thin tail on the right or left of a distribution respectively.

Test sites - those sites being tested for biological impairment.

Trophic level - a broad class of an ecosystem (e.g., green plants, herbivores, carnivores) in which all organisms procure food in the same general manner.

Use designations - predominant uses each State determines appropriate for a particular estuary, region, or area within the class.

Zooplankton - free-floating or drifting animals with movements determined by the motion of the water.

Literature Cited

Alden, R.W., S.B. Weisberg, J.A. Ranasinghe, and D.M. Dauer. 1997. Optimizing temporal sampling strategies for benthic environmental monitoring programs. *Marine Pollution Bulletin* 34(11):913-922.

Allen, M.J. and R.W. Smith. 2000. *Development of demersal fish biointegrity indices for coastal Southern California*. Southern California Coastal Water Research Project, Westminster, CA.

Allen, M.J., D. Diener, J. Mubarak, S.B. Weisberg, and S.L. Moore. 1999. Megabenthic invertebrate assemblages of the mainland shelf of southern California in 1994. Pages 113-124 *in* Weisberg, S.B. and D. Hallock (editors). *Southern California Coastal Water Research Project, Annual Report 1997-1998*. Reynolds and Reynolds, Santa Ana, CA.

Allen, H.E., G. Fu and B. Deng. 1993. Analysis of acid-volatile sulfide (AVS) and simultaneously extracted metals (SEM) for the estimation of potential toxicity in aquatic sediments. *Environmental Toxicology and Chemistry* 12:1441-1453.

Alve, E. 1991. Foraminifera, climatic change, and pollution: A study of late Holocene sediments in Drammensfjord, southeast Norway. *The Holocene* 1(3):243-261.

American Public Health Association (APHA). 1992. *Standard methods for the examination of waters and wastewater*. American Public Health Association, American Water Works Association, and Water Pollution Control Federation. 18th edition, Washington, DC.

American Public Health Association (APHA). 1981. *Standard methods for the examination of water and wastewater*. American Public Health Association, American Water Works Association, and Water Pollution Control Federation. 15th edition, Washington, DC.

American Society for Testing and Materials (ASTM). 1998a. Standard guide for conducting 10-day static sediment toxicity tests with marine and estuarine amphipods. E1367-92. Volume 11.05:732-757. *Annual Book of Standards: American Society of Testing and Materials*, Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1998b. Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. E1391-94. Volume 11.05:768-788. *Annual Book of Standards: American Society of Testing and Materials*, Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1998c. Standard guide for conducting sediment toxicity tests with marine and estuarine polychaetous annelids. E1611-94. Volume 11.05:1009-1032. *Annual Book of Standards: American Society of Testing and Materials*, Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1991. Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. ASTM Designation E1391-90. *Annual Book of Standards*. American Society for Testing and Materials, Philadelphia, PA.

Ankley, G.T., D.M. DiToro, D.J. Hansen and W.J. Berry. 1996. Technical basis and proposal for deriving sediment quality criteria for metals. *Environmental Toxicology and Chemistry* 15(12):2056-2066.

Ankley, G., M. Schubauer-Berigan, J. Dierkes. 1991. Predicting the toxicity of bulk sediments to aquatic organisms with aqueous test fractions: pore water vs. elutriate. *Environmental Toxicology and Chemistry* 10:925-939.

Association of Bay Area Governments (ABAG). 1991. *Status and trends report on wetlands and related habitats in the San Francisco estuary.* San Francisco Estuary Project. Oakland, CA.

Barbour, M.T., J. Gerritsen, B.D. Snyder, J.B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish, 2nd ed.* EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Barbour, M.T., J. Gerritsen, and J.S. White. 1996a. *Development of the Stream Condition Index (SCI) for Florida.* Prepared for FL DEP, Tallahassee, FL.

Barbour, M.T., J.M. Diamond, C.O. Yoder. 1996b. Biological assessment strategies: Applications and limitations. Pages 245-270 *in* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins (editors). *Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts.* SETAC Press, Pensacola, Florida.

Barbour, M.T., J.B. Stribling, and J.R. Karr. 1995. The multimetric approach for establishing biocriteria and measuring biological condition. *In* W. Davis, T. Simon (editors). *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making.* Lewis Publishers. Boca Raton, FL.

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. *Environmental Toxicology and Chemistry* 11:437-449.

Barss, M.S. and G.L. Williams. 1973. Palynology and nannofossil processing techniques. *Geological Survey of Canaca Paper.* 73-26, 1-25.

Batiuk, R.A., R.J. Orth, K.A. Moore, W.C. Dennison, J.C. Stevenson, L.W. Staver, V. Carter, N.B. Rybicki, R.E. Hickman, S. Kollar, S. Bieber, P. Heasly. 1992. *Chesapeake Bay Submerged Aquatic Vegetation Habitat Requirements and Restoration Targets: A Technical Synthesis.* Chesapeake Bay Program, 68-WO-0043.

Bergen, M., S.B. Weisberg, R.W. Smith, D. Cadien, A. Dalkey, D. Montagne, J.K. Stull, R.G. Velarde. 1999. Relationship between depth, latitude, and sediment and the structure of benthic infaunal assemblages on the mainland shelf of southern California. Pages 125-136 *in* Weisberg, S.B. and D. Hallock (editors). *Southern California Coastal Water Research Project, Annual Report 1997-1998.* Reynolds and Reynolds, Santa Ana, CA.

Bernstein, B.B., B.E Thompson, and R.W. Smith. 1991. *A combined science and management framework for developing regional monitoring objectives.* Presented at the National Estuary Program Science Symposium, Sarasota, FL. 25-27 February 1991.

Bilyard, G.R. 1987. The value of benthic infauna in marine pollution monitoring studies. *Marine Pollution Bulletin* 18:581-585.

Blalock, H.M., Sr. 1979. *Social Statistics*. Revised 2nd edition. McGraw-Hill Book Company, New York, NY.

Boesch, D.F. 1977. *Application of numerical classification in ecological investigations of water pollution*. EPA Grant No. R803599-01-1, ROAP/TASK No. 21 BEI, U.S. EPA Corvallis Environmental Research Laboratory, Newport, OR. 115 p.

Bowman, M.L., E. Dohner, and C. Dohner. 1993. *Summary of estuarine monitoring program attributes for: Chesapeake Bay benthos; Chesapeake Bay plankton; Tar/Pamlico; EMAP-Estuaries Virginian Province demonstration project; Naples Bay, Florida; San Francisco estuary and wetlands; Puget Sound ambient monitoring program; and Puget Sound estuary program*. Prepared for U.S. Environmental Protection Agency, Assessment and Watershed Protection Division, Oceans and Coastal Protection Division, and Health and Ecological Criteria Division, Washington, DC by Tetra Tech, Inc., Owings Mills, MD.

Boyle, T.P., G.M. Smillie, J.C. Anderson, and D.R. Bieson. 1990. A sensitivity analysis of nine diversity and seven similarity indices. *Research Journal of the Water Pollution Control Federation* 62:749.

Boynton, W.R., J.H. Garber, R. Summers, and W.M. Kemp. 1995. Inputs, transformations, and transport of nitrogen and phosphorus in Chesapeake Bay and selected tributaries. *Estuaries* 18(1B):285-314.

Boynton, W.R., L. Murray, W.M. Kemp, J.D. Hagy, C. Stokes, F. Jacobs, J. Bowers, S. Souza, B. Rinsky, and J. Seibel. 1993. *Maryland's Coastal Bays: An assessment of aquatic ecosystems, pollutant loadings, and management options*. Prepared for Maryland Department of the Environment.

Boynton, W.R., W.M. Kemp, J. Garber, J.M. Barnes, L.L. Robertson, and J.L. Watts. 1988. *Chesapeake Bay water quality monitoring program ecosystems processes component*. Level 1 Report No. 5. Prepared for Maryland Department of Environment by University of Maryland Center for Environmental and Estuarine Studies.

Brenum, G. 1976. *A comparative study of benthic communities of dredged lagoons, tidal creeks, and areas of open bays in Little Assawoman, Indian River, and Rehoboth Bays, Delaware*. M.S. thesis, College of Marine Studies, University of Delaware, Newark, DE.

Brumbaugh, W., C. Ingersoll, N. Kemble, T. May, and J. Zajicek. 1994. Chemical characterization of sediments and pore water from the Upper Clark Fork River and Milltown Reservoir, Montana. *Environmental Toxicology and Chemistry* 13:1971-1973.

Brush, G.S. 1989. Rates and patterns of estuarine sediment accumulation. *Limnology and Oceanography* 34(7):1235-1246.

Bufflap, W. and H. Allen. 1995. Sediment pore water collection methods: A review. *Water Research* 29:165-177.

Bulger, A.J., B.P. Hayden, M.E. Monaco, D.M. Nelson, and M.G. McCormick-Ray. 1993. Biologically-based estuarine salinity zones derived from a multivariate analysis. *Estuaries* 16(2):311-322.

Burgess, R., K. Schweitzer, R. McKinney, D. Phelps. 1993. Contaminated marine sediments: Water column and interstitial toxic effects. *Environmental Toxicology and Chemistry* 12:127-138.

Carmichael, J.T., B.M. Richardson, M. Roberts, and S.J. Jordan. 1992. *Fish

assemblages and dissolved oxygen trends in eight Chesapeake Bay tributaries during the summers of 1989-1991: a data report. Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD.

Carriker, M.R. 1967. Ecology of estuarine benthic invertebrates: A perspective. American Association for the Advancement of Science, Washington, DC. *Estuaries* 83:442-487.

Chaloud, D.J. and D.V. Peck, editors. 1994. *Environmental Monitoring and Assessment Program: Integrated Quality Assurance Project Plan for the Surface Waters Resource Group, 1994 Activities.* EPA 600/X-91/080, Rev. 2.00. USEPA, Las Vegas, NV.

Chapman, P.M. 1996. Presentation and interpretation of sediment quality triad data. *Ecotoxicology* 5:327-339.

Chapman, P.M. 1988. Marine sediment toxicity tests. *In* J.J. Lichtenberg, F.A. Winter, C.I. Weber, and L. Fredkin (editors). *Chemical and Biological Characterization of Sludges, Sediments, Dredge Spoils, and Drilling Muds.* Philadelphia, PA: ASTM.

Chapman, P.M., R.N. Dexter, and E.R. Long. 1987. Synoptic measures of sediment contamination, toxicity, and infaunal community structure (the Sediment Quality Triad). *Marine Ecology Progress Series* 37:75-96.

Charles et al. 1994. Paleolimnological approaches to biological monitoring. Pages 233-293 *in* L.L. Loeb, and A. Spacie (editors). *Biological monitoring of aquatic systems.* CRC Press, Boca Raton, FL.

Charles, D.F. and J.P. Smol. 1994. Long-term chemical changes in lakes:

Quantitative inferences using biotic remains in the sediment record. *Advances in Chemistry* 237:1-51.

Christian, R.R. 1989. Microbial ecology and organic detritus in estuaries. *In* J.W. Day, Jr., C.A.S. Hall, W.M. Kemp, and A. Yanez-Arancibia (editors). *Estuarine Ecology.* John Wiley & Sons, New York, NY. pp. 558.

Cochran, W.G. 1963. *Sampling Techniques.* John Wiley and Sons, Inc., New York, NY.

Conley, D.J., C.L. Shelske and E.F. Stoermer. 1993. Modification of the biogeochemical cycle of silica with eutrophication. *Marine Ecological Progress Series* 101:179-192.

Conover, W.J. 1980. *Practical Nonparametric Statistics. 2nd edition.* John Wiley and Sons, Inc., New York, NY.

Conquest, L.L., S.C. Ralph, and R.J. Naiman. 1994. Implementation of large-scale stream monitoring efforts: Sampling design and data analysis issues. Pages 69-90 *in* L. Loeb and A. Spacie (editors). *Biological Monitoring of Aquatic Systems.* Lewis Publishers, Boca Raton, FL.

Cooper, S.R. 1995. Chesapeake Bay watershed historical land use: Impact on water quality and diatom communities. *Ecological Applications* 5:703-723.

Cooper, S.R. and G.S. Brush. 1991. Long-term history of Chesapeake Bay anoxia. *Science* 254:992-996.

Correll, D.L. 1987. Nutrients in Chesapeake Bay. Pages 298-320 *in* S.K. Majumdar, L.W. Hall, Jr. and H.M. Austin (editors). *Contaminant Problems and Management of Living Chesapeake Bay*

Resources. Pennsylvania Academy of Science.

Costanza, R., S.O. Funtowicz, and J.R. Ravetz. 1992. Assessing and communicating data quality in policy-relevant research. *Environmental Management* 16(1):121-131.

Dale, B., T.A. Thorsen and A. Fjellsa. 1999. Dinoflagellate cysts as indicators of cultural eutrophication in the Oslofjord, Norway. *Estuarine, Coastal and Shelf Science* 48:371-382.

Dardeau, M.R., R.F. Madlin, W.W. Schroeder, and J.P. Stout. 1992. *Estuaries: Biodiversity of the southeastern United States.* C.T. Hackney, S.M. Adams, and W.H. Martin (editors). John Wiley & Sons, New York, NY. pp. 779.

Dauer, D.M. 1993. Biological criteria, environmental health and estuarine macrobenthic community structure. *Marine Pollution Bulletin* 26:249-257.

Davies, S.P., L.T. Somides, D.L. Courtemanch, and F. Drummond. 1993. *Maine Biological Monitoring and Biocriteria Development Program.* Maine Department of Environmental Protection, Bureau of Water Quality Control, Division of Environmental Evaluation and Lake Studies, Augusta, ME.

Day, J.W., C.A.S. Hall, W.M. Kemp, and A. Yanez-Aranciba. 1989. *Estuarine Ecology.* John Wiley & Sons, New York, NY. 558 pp.

Deegan, L.A., J.T. Finn, S.G. Ayvazian, C.A. Ryder-Kieffer, and J. Buonaccorsi. 1997. Development and Validation of an Estuarine Biotic Integrity Index. *Estuaries* 20:601-617.

Delaware Chamber of Commerce, personal communication, 1990.

Dennison, W.C., R.J. Orth, K.A. Moore, J.C. Stevenson, V. Carter, S. Kollar, P. Bergstrom, and R.A. Batiuk. 1993. Assessing water quality with submerged aquatic vegetation. *Bioscience* 43:86-94.

Diaz, R. and W. Nelson. 1993. *Middle and southern Atlantic coast estuarine benthic invertebrate metrics development.* A Proceedings: Estuarine and Near Coastal Bioassessment and Biocriteria Workshop, (November 18-19, 1992.) Annapolis, Maryland. U.S. Environmental Protection Agency, Office of Science and Technology. Washington, DC.

Dickerson, R. 1995. Atmospheric nitrogen deposition to the Chesapeake Bay. *In* P. Hill and S. Nelson (editors). *Toward a Sustainable Coastal Watershed: The Chesapeake Experiment. Proceedings of a Conference, 1-3 June 1994, Norfolk, VA.* Chesapeake Research Consortium, Edgewater, MD. CRC Publication No. 149. p. 507.

DiToro, D.M., J.D. Mahony, D.J. Hansen, K.J. Scott, M.B. Hinks, S.M. Mayr and M.S. Redmond. 1990. Toxicity of cadmium in sediments: The role of acid volatile sulfides. *Environmental Toxicology and Chemistry* 9:1487-1502.

DiToro, D.M., C. Zarba, D.J. Hansen, R.C. Swartz, C.E. Cowan, H.E. Allen, N.A. Thomas, P.R. Paquin and W.J. Berry. 1991. Technical basis for establishing sediment quality criteria for non-ionic organic chemicals using equilibrium partitioning. *Environmental Toxicology and Chemistry* 10:1299-1307.

Dixit, S.S., J.P. Smol, J.C. Kingston, D.F. Charles. 1992. Diatoms: powerful indicators of environmental change. *Environmental Science and Technology* 26(1):23-32.

Dossis, P. and L.J. Warren. 1981. Zinc and lead in background and contaminated sediments from Spencer Gulf, South Australia. *Environmental Science and Technology* 15:1451-6.

Dycus, D.L. 1995. *Aquatic ecological health determinations for TVA reservoirs - 1994. An informal summary of 1994 vital monitoring results and ecological health determination methods.* Tennessee Valley Authority Water Management.

Dycus, D.L. and D.L. Meinert. 1993. *Monitoring and evaluation of aquatic resource health and use suitability in Tennessee Valley Authority reservoirs.* Water Management. Draft.

Eaton, C.M. 1997. *Sediment wet-sieving technique to determine "percent fines."* C.M. Eaton, Bio-Marine Enterprises, Seattle, WA. Fax transmission to M. Bowman, Tetra Tech, Inc., Owings Mills, MD. 4 February 1997.

Eaton, C. M. 1995. Population patterns of demersal fauna and environmental stress: a preliminary, trawl-based assessment. *Puget Sound Notes* 36:1-6.

Eaton, C.M. 1994. *Development of trawl-based tools for the quantitative assessment of demersal fauna (macroinvertebrates and fishes): A summary of phase I and II.* Final Report to USEPA, Washington, DC. Order No. 4642.

Eaton, C.M. and P.A. Dinnel. 1994. *Development of trawl-based criteria for the assessment of demersal fauna (macroinvertebrates and fishes): Pilot study in Puget Sound, Washington.* Presented at Estuarine and Near Coastal Marine Bioassessment/ Biocriteria Workgroup Meeting. USEPA, Baltimore, MD. January 6th.

Eaton, C.M. and P.A. Dinnel. 1993. *Development of trawl-based criteria for assessment of demersal fauna (macroinvertebrates and fishes): Pilot Study in Puget Sound, Washington.* Final report to USEPA, Washington, DC.

Eaton, L. 1994. *Results of a test of three Methods, February 7, 1994.* North Carolina Division of Environmental Management Biological Assessment Group. Memorandum to Ken Eagleson, May 10, 1994.

Eaton, L. 1992a. *Pamlico basin sampling: More metal hotspots.* Memorandum dated March 11, 1992 to T. MacPherson, North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC.

Eaton, L. 1992b. *Diaz method testing results.* Memorandum date March 30, 1992 to Ken Eagleson, North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC.

Eaton, L. 1992c. *Biological results from sediment toxicity survey, Neuse and Pamlico estuaries, January 9-10, 1992. Memorandum dated April 14, 1992 to Harold Quidley, North Carolina.* Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC.

Eaton, L. 1992d. Letter dated September 9, 1992 from L. Eaton, North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC, to M. Bowman, Tetra Tech, Inc., Owings Mills, MD.

Engle, V.D. and J.K. Summers. 1999. Refinement, validation, and application of a benthic condition index for northern

Gulf of Mexico estuaries. *Estuaries* 22:624-635.

Engle, V.D., J.K. Summers, and G.R. Gaston. 1994. A benthic index of environmental condition of Gulf of Mexico estuaries. *Estuaries* 17(2):372-384.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine Freshwater Research* 42:555-67.

Farrell, D.H. 1993a. *A community based metric for marine benthos (Draft).* Florida Department of Environmental Protection, Tampa, FL.

Farrell, D. H. 1993b. *Bioassessment in Florida.* Pages 17-26 *in* A Proceedings: Estuarine and Near Coastal Bioassessment and Biocriteria Workshop, Annapolis, MD. USEPA, Office of Science and Technology, Washington, D.C.

Ferguson, R.L. and L.L. Wood. 1994. *Rooted vascular aquatic beds in the Albemarle-Pamlico estuarine system.* National Marine Fisheries Service, Beaufort, NC. Project No. 94-02.

Ferraro, S.P. and F.A. Cole. 1990. Taxonomic level and sample site sufficient for assessing pollution impacts on the southern California macrobenthos. *Marine Ecology Progress Series* 67:251-262.

Ferraro, S.P. and F.A. Cole. 1992. Taxonomic level sufficient for assessing a moderate impact on macrobenthic communities in Puget Sound, Washington, D.C. *Canadian Journal of Fisheries and Aquatic Sciences* 49(b):1184-1188.

Ferraro, S.P. and F.A. Cole. 1995. Taxonomic level sufficient for assessing pollution impacts on the southern California bight macrobenthos - revisited. *Environmental Toxicology and Chemistry* 14(6):1031-1040.

Ferraro, S.P., F.A. Cole, W.A. DeBen, and R.C. Swartz. 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, USA. *Canadian Journal of Fisheries and Aquatic Sciences* 46(10):2157-2165.

Ferraro, S.P., R.C. Swartz, F.A. Cole, and W.A. DeBen. 1994. Optimum macrobenthic sampling protocol for detecting pollution impacts in the southern California Bight. *Environmental Monitoring and Assessment* (29):127-153.

Flannagan, J.F. 1970. Efficiencies of various grabs and corers in sampling freshwater benthos. *Journal of the Fisheries Research Board of Canada* 27:1631-1700.

Flint, R.W. and R.D. Kalke. 1985. Benthos Structure and function in a south Texas estuary. *Contributions in Marine Science* 28:33-53.

Forstner, U. and G.T.W. Wittmann. 1981. *Metal pollution in the aquatic environment.* Second revised edition. Springer-Verlag, New York, NY.

Fredette, T.J., D.A. Nelson, T. Miller-Way, J.A. Adair, V.A. Sotler, J.E. Clausner, E.B. Hands, and F.J. Anders. 1989. *Selected tools and techniques for physical and biological monitoring of aquatic dredged material disposal sites.* Final Report. U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Frydenborg R. 1994. *Lake bioassessments for the determination of nonpoint source impairment in Florida.* Draft. FL DEP, Biology Section, Division of Administrative and Technical Services, Tallahassee, FL. July.

Gaston, G.R., D.L. Lee, and J.C. Nasci. 1988. Estuarine macrobenthos in Calcasieu Lake, Louisiana: Community and trophic structure. *Estuaries* 11:192-200.

Gibson, G.R. 1995. *Near coastal marine waters pilot project* (unpublished report). USEPA, Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC.

Gibson, G.R. et al. 1993. *Proceedings Estuarine and Near Coastal Marine Bioassessment and Biocriteria Workshop, 18-19.* November, 1992.

Gibson, G.R. 1992. *Procedures for initiating narrative biological criteria.* EPA-822-B-92-002. USEPA, Office of Science and Technology, Washington, DC.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* Van Nostrand Reinhold, New York. 320pp.

Glew, J.R. 1988. A portable extruding device for close interval sectioning of unconsolidated core samples. *Journal of Paleolimnology* 1:235-239.

Goodyear, C.P. 1985. Relationship between reported commercial landings and abundance of striped bass in Chesapeake Bay, Maryland. *Transactions of the American Fisheries Society* 114(1):92-96.

Gray, J.S. 1989. Effects of environmental stress on species rich assemblages. *Biological Journal of the Linnean Society* 37:19-32.

Green, R.H. 1984. Some guidelines for the design of biological monitoring programs in the marine environment. Pages 233-245 *in* H.H. White (editor). *Concepts of Marine Pollution Measurements.* University of Maryland Sea Grant, College Park, MD.

Guillen, G. 1995a. *Development of a rapid bioassessment method and index of biotic integrity in southeast Texas.* Presented at Estuarine and Near Coastal Marine Bioassessment/Biocriteria Workgroup Meeting. U.S. EPA, Baltimore, MD. January 6th.

Guillen, G.J. 1995b. *Development of a Rapid Bioassessment method and index of biotic integrity for coastal environments located along the northwest Gulf of Mexico - DRAFT.* Texas National Resource Conservation Commission, Environmental Assessment Program, Field Operations Division.

Guillen, G. 1994. *Development of a rapid bioassessment method and index of biotic integrity in southeast Texas.* Presented at Estuarine and Near Coastal Marine Bioassessment/Biocriteria Workgroup Meeting. USEPA, Baltimore, MD.

Hansen, D.J., Berry, W.J., Mahony, J.D., Boothman, W.S., DiToro, D.M., Robson, D.L., Ankley, G.T., Ma, D., Yan, Q., Pesch, C.E. 1996. Predicting the toxicity of metal-contaminated field sediments using interstitial concentration of metals and acid-volatile sulfide normalizations. *Environmental Toxicology and Chemistry* 15:(12) 2080-2094.

Helsel, D.R. and R.M. Hirsch. 1992. *Statistical methods in water resources.* Elsevier, Amsterdam, Netherlands.

Hillman, K., D.I. Walker, A.W.D. Larkum, and A.J. McComb. 1989. Productivity and nutrient limitation. Pages 635-685 *in* A.W.D. Larkum, A.J.

McComb, and S.A. Shepard (editors). *Biology of Seagrasses. A treatise on the biology of seagrasses with special reference to the Australian region.* Elsevier, New York, NY.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomologist* 20:31-39.

Hinga, K.R. 1988. Seasonal predictions for pollutant scavenging in two coastal environments using a model calibration based upon thorium scavenging. *Marine Environmental Research* 26:97-112.

Holland, A.F. (editor). 1990. *Near coastal program plan for 1990: Estuaries.* EPA/600/4-90/033. Office of Research and Development, USEPA, Narragansett, RI.

Holland, A.F. 1985. Long-term variation of macrobenthos in a mesohaline region of the Chesapeake Bay. *Estuaries* 8(2a):93-113.

Holland, A.F., A.T. Shaughnessy, L.C. Scott, V.A. Dickens, J. Gerritsen, and J.A. Ranasinghe. 1989. *Long-term benthic monitoring and assessment program for the Maryland portion of the Chesapeake Bay: Interpretive report.* CBRM-LTB/EST-2. Prepared for MDNR, Power Plant Research Program, Annapolis, Maryland by Versar, Inc., Columbia, MD.

Holland, A.F., A.T. Shaughnessy, L.C. Scott, V.A. Dickens, J.A. Ranasinghe, and J.K. Summers. 1988. *Progress report: Long-term benthic monitoring and assessment program of the Maryland portion of Chesapeake Bay (July 1986 - October 1987).* PPRP-LTB/EST-88-1. Prepared for MDNR, Power Plant Research Program, Annapolis, Maryland by Versar, Inc., Columbia, MD.

Holland, A.F., A.T. Shaughnessy, and M.H. Hiegel. 1987. Long-term variation in mesohaline Chesapeake Bay macrobenthos: spatial and temporal patterns. *Estuaries* 10(3):227-245.

Honeyman, B.D. and P.H. Santschi. 1988. Metals in aquatic systems: Predicting their scavenging residence times from laboratory data remains a challenge. *Environmental Science and Technology* 22:862-871.

Houde, E.D, and Zastrow, C.E. 1991. Bay Anchovy *Anchoa mitchilli*. Pages 81-86 *in* S.L. Funderbunk, J.A. Mihursky, S.J. Jordan, and D. Riley (editors). *Habitat requirements for Chesapeake Bay living resources.* Prepared for the Living Resources Subcommittee, Chesapeake Bay Program.

Howard, R.K., G.J. Edgar, and P.A. Hutchings. 1989. Faunal assemblages of seagrass beds. Pages 536-564 *in* A.W.D. Larkum, A.J. McComb, and S.A. Shepard (editors). *Biology of seagrasses. A treatise on the biology of seagrasses with special reference to the Australian region.* Elsevier, New York, NY.

Hughes, R.M. and D.P. Larsen. 1988. Ecoregions: An approach to surface water protection. *Journal of the Water Pollution Control Federation* 60:486-493.

Hughes, R.M., P.R. Kaufmann, A.T. Herlihy, T.M. Kincaid, L. Reynolds and D.P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Science* 55:1618-1631.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: A method for assessing stream potentials. *Environmental Management* 10:629-635.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field

experiments. *Ecological Monographs* 54(2).

Hutchinson, G.E. 1967. *A treatise on limnology*. Wiley, New York, NY.

Hyland, J.L., L. Balthis, C.T. Hackney, G. McRae, A.H. Ringwood, T.R. Snoots, R.F. Van Dolah, and T.L. Wade. 1998. *Environmental quality of estuaries of the Carolinian Province: 1995*. NOAA Technical Memorandum NOS ORCA 123. National Oceanic and Atmospheric Administration, Charleston, SC.

Hyland, J.L., T.J. Herrlinger, T.R. Snoots, A.H. Ringwood, R.F. Van Dolah, C.T. Hackney, G.A. Nelson, J.S. Rosen, and S.A. Kokkinakis. 1996. *Environmental quality of estuaries of the Carolinian Province: 1994*. Annual Statistical Summary for the 1994 EMAP-Estuaries Demonstration Project in the Carolinian Province. NOAA Technical Memorandum NOS ORCA 97. National Oceanic and Atmospheric Administration/NOS, Office of Ocean Resources Conservation and Assessment, Silver Spring, MD. 102 p.

Ingersoll, C.G., P.S. Haverland, E.L. Brunson, T.J. Canfield, F.J. Dwyer, C.E. Henke, N.E. Kemble. 1996. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius*. *Journal of Great Lakes Research* 22(3): 602-623.

Jordan, S., J. Carmichael, and B. Richardson. 1992. *Habitat measurements and index of biotic integrity based on fish sampling in northern Chesapeake Bay*. A Proceedings: Estuarine and Near Coastal Bioassessment and Biocriteria Workshop, Annapolis, Maryland. USEPA, Office of Science and Technology, Washington, DC.

Kajak, Z. 1963. Analysis of quantitative benthic methods. *Ekologia Polska* (A) 11:1-56.

Karr, J.R. 1991. Biological integrity: A long-neglected aspect of water resource management. *Ecological Applications* 1:66-84.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6(6):21-27.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. *Assessing biological integrity in running waters: A method and its rationale*. Illinois Natural History Survey, Spec. Publ. 5.

Kemble, N., W. Brumbaugh, E. Brenson, F. Dwyer, C. Ingersoll, D. Monda, and D. Woodward. 1994. Toxicity of metal contaminated sediments from the Upper Clark Fork River Montana to aquatic invertebrates in laboratory exposures. *Environmental Toxicology and Chemistry* 13:1985-1997.

Kendall, D.K. 1983. *The role of physical-chemical factors in structuring subtidal marine and estuarine benthos*. Tech. Rep. EL-83-2. U.S. Army Waterways Exp. Stn., Vicksburg, MS.

Klemm, D.J., G.J. Strober, and J.M. Lazorchak. 1992. *Fish field and laboratory methods for evaluating the biological integrity of surface waters*. EPA/600/R-92-111. USEPA, Cincinnati, OH.

Komar, P.D. 1976. *Beach processes and sedimentation*. Prentice-Hall, Inc. Englewood Cliffs, NJ. 429 pp.

Krom, M.D. and R.A Berner. 1983. *Journal of Sedimentary Petrology* 53, 660.

Krumgalz, B.S. 1993. "Fingerprints" approach to the identification of anthropogenic trace metal sources in the

nearshore and estuarine environments. *Estuaries* 16(3A):488-495.

LaPointe, B.E. and M.W. Clark. 1992. Nutrient inputs from the watershed and coastal eutrophication in the Florida keys. *Estuaries* 15(4):465-476.

Latimer, J., W. Boothman, R. Tobin, D. Keith, J. Kiddon, D. Scott, S. Jayaraman, R. McKinney, and G. Chmura. 1997. Historical reconstruction of contaminant levels and ecological effects in a highly contaminated estuary. *14th International Conference of the Estuarine Research Federation.* Providence, RI.

Lenat, D.R. 1993. A biotic index for the southeastern United States: Derivation and list of tolerance values, with criteria for assigning water quality ratings. *Journal of the North American Benthological Society* 12(3):279-290.

Lie, U. 1974. Distribution and structure of benthic assemblages in Puget Sound, Washington, USA. *Marine Biology* 26:203-223.

Llansó, R.J. 1999. The distribution and structure of soft-bottom macrobenthos in Puget Sound in relation to natural and anthropogenic factors. *Puget Sound Research 1998.* Puget Sound Ambient Monitoring Program, Olympia, WA: 760-771.

Long, E.R. 1989. The use of the sediment quality triad in classification of sediment contamination. *Marine Board, National Research Council Symposium/Workshop on Contaminated Marine Sediments.* National Research Council, Washington, D.C.

Long, E.R., L.J. Field, and D.D. MacDonald. 1998a. Predicting toxicity in marine sediments with numerical sediment quality guidelines.

Environmental Toxicology and Chemistry 17(4):714-727.

Long, E.R., D. D. MacDonald, J.C. Cubbage, and C.G. Ingersoll. 1998b. Predicting the toxicity of sediment-associated trace metals with SEM:AVS concentrations and dry weight-normalized concentrations: A critical comparison. *Environmental Toxicology and Chemistry* 17(4):972-974.

Long, E.R., A. Robertson, D.A. Wolfe, J.Hameedi, and G.M. Sloane. 1996. Estimates of the spatial extent of sediment toxicity in major U.S. estuaries. *Environmental Science & Technology* 30(12):3585-3592.

Long, E.R., D.D. MacDonald, S.L. Smith, F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* 19:81-97.

Long, E.R. and L.G. Morgan. 1990. *The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program.* NOAA Technical Memorandum NOSOMA52. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Rockville, MD.

Long, E., M. Buchman, S. Bay, R. Breteler, R. Carr, P. Chapman, J. Hose, A. Lissner, J. Scott, and D. Wolfe. 1990. Comparative evaluation of five toxicity tests with sediments from San Francisco Bay and Tomales Bay, CA. *Environmental Toxicology and Chemistry* 9:1193-1214.

Long, E.R. and P.M. Chapman. 1985. A sediment quality triad: Measures of sediment contamination, toxicity and infaunal community composition in

Puget Sound. *Marine Pollution Bulletin* 16:405-415.

Luckenback, M.W., R.J. Diaz, and L.C. Schaffner. 1988. *Benthic assessment procedures.*

Ludwig, J.A. and J.F. Reynolds. 1988. *Statistical ecology. A primer on methods and computing.* 337 p. John Wiley and Sons, New York, NY.

Mallin, M.A. 1994. Phytoplankton ecology of North Carolina estuaries. *Estuaries* 17(3):561-574.

Malone, T.C., D.J. Conley, T.R. Fisher, P.M. Gilbert, L.W. Harding, and K.G. Sellner. 1996. Scales of nutrient-limited phytoplankton productivity in Chesapeake Bay. *Estuaries* 19(2B):371-385.

Maryland Chamber of Commerce, personal communication, 1990.

MacDonald, D.D., R.S. Carr, F.D. Calder, E.R. Long, and C.G. Ingersoll. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253-278.

Mearns, A.J. and J.Q. Word. 1982. Forecasting effects of sewage solids on marine benthic communities. *In* G. F. Mayer (editor). *Ecological stress and the New York Bight: Science and management.* Estuarine Research Federation. 713 pp.

Miller, D.L., P.M. Leonard, R.M. Hughes, J.R. Karr, P.B. Moyle, L.H. Schrader, B.A. Thompson, R.A. Daniels, K.D. Fausch, G.A. Fitzhugh, J.R. Gammon, D.B. Halliwell, P.L. Angermeier, and D.J. Orth. 1988. Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13:12-20.

National Research Council (NRC). 1989. *Contaminated Marine Sediments – Assessment and Remediation.* National Academy Press, Washington, DC.

Nelson, W.G. 1990. Prospects for development of an index of biotic integrity for evaluating habitat degradation in coastal systems. *Chemistry and Ecology* 23: 152-165.

Nelson, W.G., and F.D. Spoon. 1994a. *Field verification of marine metrics developed for benthic habitats.* Florida Institute of Technology, Oceanography Program, Division of Marine and Environmental Systems. Final Report to Tetra Tech, Inc., Owings Mills, MD.

Nelson, W.G., and F.D. Spoon. 1994b. *Field Verification of the Use of Amphipods As Bioindicators in Marine Coastal Ecosystems.* Florida Institute of Technology, Oceanography Program, Division of Marine and Environmental Systems. Final Report to Tetra Tech, Inc., Owings Mills, MD.

Nelson, W.G., R.J. Diaz, and F.D. Spoon. 1993. *Comparison of results of marine benthic metric development between Chesapeake Bay and the Indian River Lagoon, Florida.* Florida Institute of Technology, Division of Marine and Environmental Systems, Final Report to Tetra Tech, Inc., Owings Mills, MD.

Nelson, M., P. Landrum, G. Burton, Jr., J. Klaine, E. Crecelius, T. Byl, D. Gossiaux, V. Tsymbal, L. Cleveland, C. Ingersoll, and G. Sasson-Brickson. 1993. Toxicity of contaminated sediments in dilution series with control sediments. *Chemosphere* 27:1789-1812.

Nixon, S.W., C.D. Hunt and B.L. Nowicki. 1986. The retention of nutrients (C, N, P), heavy metals (Mn, Cd, Pb, Cu), and petroleum hydrocarbons in Narragansett Bay.

Pages 99-122 *in* P. Lasserre and J.M. Martin (editors). *Biogeochemical Processes at the Land-Sea Boundary.* Elsevier, New York, NY.

O'Connor, T.P, A.Y. Cantillo, and G.G. Lauenstein. 1994. Monitoring of temporal trends in chemical contamination by the NOAA National Status and Trends Mussel Watch Project. Pages 29-50 *in* K.J.M. Kramer (editor). *Biomonitoring of Coastal Waters and Estuaries.* CRC Press, Inc., Boca Raton, FL.

Odum, E.P. 1971. *Fundamentals of ecology, 3rd Edition.* W.B. Saunders Co. Philadelphia, PA.

Odum, W. E. 1970. Insidious alteration of the estuarine environment. *Transactions of the American Fisheries Society* 4:836-850.

Odum, W.E. and C.C. McIvor. 1990. Mangroves. Chapter 15 *in* Myers, R.L. and J.J. Ewel (editors). *Ecosystems of Florida.* University of Central Florida Press.

Office of Technology Assessment (OTA). 1987. *Wastes in marine environments.* OTA, Washington, DC.

Ohio EPA. 1990. *The use of biocriteria in the Ohio EPA surface water monitoring and assessment program.* Ecological Assessment Section, Division of Water Quality Planning and Assessments, Ohio EPA, Columbus, OH. August 22.

Ohio EPA. 1987. *Biological criteria for the protection of aquatic life.* Vols. 1-3. Monitor. Assess. Prog., Surface Water Sec., Div. Water Quality, Ohio EPA, Columbus, OH.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77(1):118-125.

Oreskes, N., K. Schrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641-646.

Orth, R.J., J.F. Nowak, G.F. Anderson, and J.R. Whiting. 1993. *Distribution of submerged aquatic vegetation in the Chesapeake Bay and its tributaries and Chincoteague Bay - 1992.* Prepared by Virginia Institute of Marine Science, Gloucester Point, VA for the USEPA, Chesapeake Bay Program Office, Annapolis, MD.

Overton, W.S., D. White, and D.L. Stevens. 1990. *Design report for EMAP: Environmental Monitoring Assessment Program.* EPA/600/3-91/053. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

Pait, A.S., D.R.G. Farrow, J.A. Lowe, and P.A. Pacheis. 1989. *Agricultural pesticide use in estuarine drainage areas: A preliminary summary for selected pesticides.* Strategic Assessment Branch Office of Oceanography and Marine Assessment, NOAA, Rockville, MD.

Patterson, R.T. 1990. Intertidal benthic foraminiferal biofaces on the Fraser river delta - British Columbia - Modern distribution and paleoecological importance. *Micropaleontology* 36:229-245.

Paul, J.F., J.H. Gentile, K.J. Scott, S.C. Schimmel, D.E. Campbell, and R.W. Latimer. 1999. *EMAP-Virginian Province Four-Year Assessment Report (1990-1993).* EPA 600/R-99/004. U.S. Environmental Protection Agency, Atlantic Ecology Division, Narragansett, RI.

Paulsen, S.B., D.P. Larsen, P.R. Kaufmann, T.R. Whittier, J.R. Baker, D.V. Peck, J. McGue, R.M. Hughes, D. McMullen, D.L. Stevens, J.L. Stoddard, J. Lazorchak, W. Kinney, A.R. Selle, and R. Hjort. 1991. *The Environmental Monitoring and Assessment Program (EMAP) - Surface waters monitoring research plan, fiscal year 1991.* EPA/600/3-91/022. USEPA, Corvallis, OR.

Pearson, T.H. and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology Annual Review* 16:229-311.

Peterman, R.M. 1990. The importance of reporting statistical power: The forest decline and acidic deposition example. *Ecology* 71(5):2024-2027.

Peters, J.A. 1988. Quality control infusion into stationary source sampling. *In* L.H. Keith (editor). *Principles of environmental sampling.* American Chemical Society, Washington, DC.

Peters, R.H. 1991. *A critique for ecology.* Cambridge University Press, Cambridge, MA.

Pielou, E.C. 1977. *Mathematical ecology.* John Wiley & Sons, Inc., New York, NY. 385 pp.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. *Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish.* EPA/440/4-89-001. USEPA, Office of Water, Washington, DC.

Plumb, R.H. 1981. *Procedure for handling and chemical analysis of sediment and water samples.* Technical Report EPA/CE-81-1. Prepared for the U.S. Environmental Protection Agency/Corps of Engineers Technical Committee on Criteria for Dredge and Fill Material. Environmental Laboratory, U.S. Army Waterways Experiment Station, Vicksburg, MS.

Pond, S. and G.L. Pickard. 1983. *Introductory dynamic oceanography.* 3rd Edition. Pergamon Press, Inc., New York, NY.

Puget Sound Estuary Program (PSEP). 1995. *Recommended guidelines for conducting laboratory bioassays on Puget Sound sediments.* Prepared for USEPA Region 10, Office of Puget-Sound, Seattle, WA and Puget Sound Water Quality Authority, Olympia, WA.

Puget Sound Water Quality Authority (PSWQA). 1991. *Puget Sound update.* Second annual report of the Puget Sound Ambient Monitoring Program. PSWQA, Seattle, WA. 99 pp.

Puget Sound Water Quality Authority (PSWQA). 1990. *Puget Sound update.* First annual report of the Puget Sound Ambient Monitoring Program. PSWQA, Seattle, WA. 89 pp.

Puget Sound Water Quality Authority (PSWQA). 1988. *Puget Sound ambient monitoring program, monitoring management committee.* Final Report. PSWQA, Seattle, WA. 145 pp.

Rabalais, N.N. 1990. Biological communities of the south Texas continental shelf. *American Zoologist* 30:77-87.

Rabalais, N.N., M.J. Dagg, and D.F. Boesch. 1985. *Nationwide review of oxygen depletion and eutrophication in estuarine and coastal waters: Gulf of Mexico (Alabama, Mississippi, Louisiana, and Texas).* Final Report to U.S. Department of Commerce, NOAA,

Ocean Assessment Division, Rockville, MD.

Rakoncinski, C.F., S.S. Brown, G.R. Gaston, R.W. Heard, W.W. Walker, and J.K. Summers. 1997. Macrobenthic responses to natural and contaminant-related gradients in northern Gulf of Mexico estuaries. *Ecological Applications* 7:1278-1298.

Rakocinski, C., R.W. Heard, T. Simons, and D. Gledhill. 1991. Macroinvertebrate associations from beaches of selected barrier islands in the Northern Gulf of Mexico: Important environmental relationships. *Bulletin of Marine Science* 48:689-701.

Ranasinghe, J.A., D.M. Dauer, L.C. Schaffnter, R.J. Diaz. 1994. *Chesapeake Bay Benthic Community Restoration Goals.* Chesapeake Bay Program Office, 68-D9-0166; Chesapeake Bay Research & Monitoring Division, CB92-006-004.

Ranasinghe, J.A., L.C. Scott, and R. Newport. 1992. *Chesapeake Bay water quality monitoring program: long-term benthic monitoring and assessment component.* Vol. I. Draft. Prepared by Versar, Inc., Columbia, Maryland, for the Maryland Department of Environment, Baltimore, MD.

Reckhow, K.H. and W. Warren-Hicks. 1996. *Biological criteria: Technical guidance for survey design and statistical evaluation of biosurvey data.* Prepared by School of the Environment, Duke University, Durham, NC for USEPA, Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC.

Reid, G.K. and R.D. Wood. 1976. *Ecology of inland waters and estuaries.* D. Van Nostrand Company, New York, NY. 485 pp.

Reish, D. J. and J. L. Bernard. 1960. Field toxicity tests in marine water utilizing the polychaetous annelid *Capitella capitata (Fabricius). Pacific Naturalist* 1:1-8.

Reishi D. and J. Lemay. 1988. *Bioassay manual for dredged materials.* US Army Corps of Engineers, Los Angeles District, Los Angeles, CA, Technical Report DACW-09-83R-005.

Reynoldson, T.B. and M.A. Zarull. 1993. An approach to the development of biological sediment guidelines. Pages 177-200 *in* S. Woodley, J. Kay, and G. Grancis (editors). *Ecological Integrity and the Management of Ecosystems.* St. Lucie Press.

Rhoads, D.C. 1974. Organism-sediment relations on the muddy sea floor. *Oceanography and Marine Biology Annual Review* 12:263-300.

Sanford, L.P., K.G. Sellner, and D.L. Breitburg. 1990. Covariability of dissolved oxygen with physical characteristics in the summertime Chesapeake Bay. *Journal of Marine Research* 48:567-590.

Santangelo, R. 1996. County Sanitation Districts of Orange County, 10844 Ellis Ave., Fountain Valley, CA 92728.

Sarda, N. and G. Burton, Jr. 1995. Ammonia variation in sediments: spatial, temporal, and method-related effects. *Environmental Toxicology and Chemistry* 14:1499-1506.

Schimmel, S.C., B.D. Melzian, D.E. Campbell, C.J. Stubel, S.J. Benyi, J.S. Rosen, and H.W. Buffum. 1994. *Statistical Summary: EMAP-Estuaries Virginian Province - 1991.* EPA/620/R-94/005. Office of Research and Development, USEPA, Narragansett, RI.

Schindler, D.W. 1988. Effects of acid rain on freshwater ecosystems. *Science* 239:149-158.

Schindler, D.W. 1971. A hypothesis to explain differences and similarities among lakes in the Experimental Lakes Area, northwestern Ontario. *Journal of the Fisheries Research Board of Canada* 28:295-301.

Schindler, D.W., S.E.M. Kasian and R.H. Hesslein. 1989. Biological impoverishment in lakes of the midwestern and northeastern United States from acid rain. *Environmental Science and Technology* 23:573-580.

Schlekat, C.E., B.L. McGee, D.M. Boward, E. Reinharz, D.J. Velinsky, and T.L. Wade. 1994. Tidal river sediments in the Washington, DC area. III. Biological effects associated with sediment contamination. *Estuaries* 17(2):334-344.

Schroeder, W.W. 1979. The dissolved oxygen puzzle of the Mobile estuary. *In* Jr. and J.P. Smith (editors). *H.A. Loyacano, Symposium on the Natural Resources of the Mobile Estuary, Alabama, May 1979.* Alabama Coastal Area Board, Mississippi-Alabama Sea Grant Consortium, U.S. Fish and Wildlife Service.

Schubel, J.R. and H.H. Carter. 1984. The estuary as a filter for fine-grained suspended sediment. Pages 81-104 *in* V.S. Kennedy (editor). *The Estuary as a Filter*. Academic Press, Orlando, FL.

Sen Gupta, B.K., R.E. Turner and N.N. Rabalais. 1996. Seasonal oxygen depletion in continental-shelf waters of Louisiana: Historical record of benthic foraminifers. *Geology* 24(3):227-230.

Sharpe, J.H., J.R. Pennock, T.M. Church, T.M. Tramontano, and L.A. Cifuentes.

1984. The estuarine interaction of nutrients, organics, and metals: A case study in the Delaware Estuary. Pages 241-258 *in* V.S. Kennedy (editor). *The estuary as a filter.* Academic Press, Orlando, FL.

Shepard, S.A., A.J. McComb, D.A. Bulthis, V. Neverauskas, D.A. Steffensen, and R. West. 1989. Decline of seagrasses. Pages 346-393 *in* A.W.D. Larkum, A.J. McComb, and S.A. Shepard (editors). *Biology of seagrasses. A treatise on the biology of seagrasses with special reference to the Australian region.* Elsevier, New York, NY.

Simenstad, C.A., C.D. Tanner, R.M. Thom, and L.L. Conquest. 1991. *Puget Sound Estuary Program: Estuarine habitat assessment protocol.* EPA 910/9-91-037. Prepared for USEPA, Region 10, Office of Puget Sound, Seattle, WA. 201 pp.

Simpson, B.L., R. Aaron, J. Betz, D. Hicks, J. van der Kreeke, and B. Yokel. 1979. *The Naples Bay Study.* Prepared for the Collier County Conservancy, Naples, FL.

Smith, S.M. and G.L. Hitchcock. 1994. Nutrient enrichments and phytoplankton growth in the surface waters of the Louisiana Bight. *Estuaries* 17(4):740-753.

Smith, R.W., M. Bergen, S.B. Weisberg, D. Cadien, A. Dalkey, D. Montagne, J.K. Stull, & R.G. Velarde. 2000. Benthic response index for assessing infaunal communities on the mainland shelf of southern California. *Ecological Applications.* In press.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical methods. 7th edition*. The Iowa State University Press, Ames, IA.

Snelgrove, P.V.R. and C.A. Butman. 1994. Animal-sediment relationships

revisited: Cause versus effect. *Oceanography and Marine Biology Annual Review* 32:111-177.

Southerland, M.T. and J.B. Stribling. 1995. Status of biological criteria development and implementation. Pages 81-96 *in* W.S. Davis and T.P. Simon (editors). *Biological assessment and criteria: Tools for water resources planning and decision making.* Lewis Publishers, Boca Raton, FL.

Stevenson, J.C., L.W. Staver, & K.W. Staver. 1993. Water quality associated with survival of submersed aquatic vegetation along an estuarine gradient. *Estuaries* 16(2)346-360.

Stickney, R.R. 1984. *Estuarine ecology of the southeastern United States and Gulf of Mexico.* Texas A & M University Press, College Station, TX.

Stoermer, E.F., J.A. Wolin, C.L. Schelske, and D.J. Conley. 1990. Siliceous microfossil succession in Lake Michigan. *Limnology and Oceanography* 35:959-967.

Strobel, C.J., H.W. Buffum, S.J. Benyi, E.A. Petrocelli, D.R. Reifsteck, and D.J. Keith. 1995. *Statistical Summary: EMAP-Estuaries Virginian Province - 1990 to 1993.* EPA/620/R-94/026. U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, RI.

Strobel, C.J., S.J. Benyi, D.J. Keith, H.W. Buffum, and E.A. Petrocelli. 1994. *Statistical summary: EMAP-Estuaries Virginian Province - 1992.* EPA/620/R-94/019. USEPA, Environmental Research Laboratory, Narragansett, RI.

Summers J.K. 1994. *(EMAP-Estuaries - A report on the condition of the estuaries of the United States in 1990-1993: A program in progress.)* Gulf Breeze (FL): U.S.

Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory. EPA/620/R-94/

Summers, J.K., T.L. Wade, V.D. Engle, and Z.A. Malneb. 1996. Normalization of metal concentrations in estuarine sediments from the Gulf of Mexico. *Estuaries* 19(3):581-594.

Summers, J.K., J.M. Macauley, P.T. Heitmuller, V.D. Engle, A.M. Adams, and G.T. Brooks. 1993. *Statistical summary: EMAP-Estuaries Louisianian Province - 1991.* EPA/600/R-93/001. USEPA, Office of Research and Development, Gulf Breeze, FL.

Summers, J.K. and V.D. Engle. 1993. Evaluation of sampling strategies to characterize dissolved oxygen conditions in northern Gulf of Mexico estuaries. *Environmental Monitoring and Assessment* 24:219-229.

Summers, J.K., J.M. Macauley, and P.T. Heitmuller. 1992. *Field activities report: Louisianian Province* USEPA, Office of Research and Development, Environmental Research Laboratory, Gulf Breeze, FL. ERL-Gulf Breeze Contribution No. SR-118.

Suter, G.W., II. 1993. *Ecological risk assessment.* Lewis Publishers, Boca Raton, FL.

Swartz, R., D. Schults, R. Ozretich, J. Lamberson, F. Cole, T. DeWitt. 1995. PAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field collected sediments. *Environmental Toxicology and Chemistry* 14:1972-1987.

Swartz, R.C., F.A. Cole, J.O. Lamberson, S.P. Ferraro, D.W. Schults, W.A. DeBen, H. Lee II, and R.J. Ozretich. 1994. Sediment toxicity, contaminations and

amphipod abundance at a DDT-and dieldrin-contaminated site in San Francisco Bay. *Environmental Toxicology and Chemistry* 13(6):949-962.

Swartz, R.C., F.A. Cole, D.W. Schults, and W.A. DeBen. 1986. Ecological changes in the Southern California Bight near a large sewage outfall: Benthic conditions in 1980 and 1983. *Marine Ecology Progress Series* 31:1-13.

Swartz, R.C., D.W. Schults, G.R. Ditsworth, W.A. DeBen, and F.A. Cole. 1985. Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. Pages 152-175 *in* T.P. Boyle (editor). *Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems, ASTM STP 865.* American Society for Testing and Materials, Philadelphia.

Symposium on the Classification of Brackish Waters. 1958. The Venice System for the Classification of Marine Waters According to Salinity. *Oikos* 9:311-312.

ter Braak, C.J.F. 1986. Canonical correspondence analysis: new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179.

Thomann, R.V. and J.A. Mueller. 1987. *Principles of surface water quality modeling and control.* Harper and Row Publishers, New York, NY. 694 pp.

Thompson, B.A., and G.R. Fitzhugh. 1986. *A use attainability study: An evaluation of fish and macroinvertebrate assemblages of the Lower Calcasieu River, Louisiana.* LSU-CFI-85-29. Louisiana State University Center for Wetland Resources, Coastal Fisheries Institute, Baton Rouge, LA.

Thompson, S.K. 1992. *Sampling.* John Wiley & Sons, Inc., New York, NY. 343 pp.

Turekian, K.K. 1977. The fate of metals in the oceans. *Geochimica et Cosmochimica Acta* 41:1139-1144.

Turner, R.E. and N.N. Rabalais. 1994. Coastal eutrophication near the Mississippi River delta. *Nature* 368:619-621.

Turner, R.E., W.W. Schroder, and W.J. Wiseman. 1987. The role of stratification in the deoxygenation of Mobile Bay and adjacent shelf bottom waters. *Estuaries* 10:13-19.

U.S. Army Corps of Engineers. 1996. *A comparison of the benthic macrofaunal resources within the Bethany Beach Sand Borrow Area.* Prepared for U.S. Army Corps of Engineers, Philadelphia District by Versar, Inc., Columbia, MD.

U.S. Environmental Protection Agency (USEPA). 1998a. *EPA requirements for quality assurance project plans for environmental data operations.* EPA QA/R-5. U.S. Environmental Protection Agency, Quality Assurance Division, Washington, DC 20460.

U.S. Environmental Protection Agency (USEPA). 1998b. *Lake and reservoir bioassessment and biocriteria technical guidance document.* U.S. Environmental Protection Agency, Office of Water, Washington, D.C. EPA-841-B-98-007.

U.S. Environmental Protection Agency (USEPA). 1996a. *Biological criteria: Technical guidance for streams and small rivers.* EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1996b. *Recommended*

guidelines for sampling and analyses in the Chesapeake Bay monitoring program. CBP/TRS 148/96. EPA 903-R-006. Chesapeake Bay Program, Annapolis, MD.

U.S. Environmental Protection Agency (USEPA). 1995. *Bibliography of methods for marine and estuarine monitoring.* EPA 842-B-95-002. USEPA, Office of Water, Office of Wetlands, Oceans, and Watersheds, Oceans and Coastal Protection Division, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1994a. *CWA Section 403: Procedural and monitoring guidance.* EPA 842-B-94-003. Office of Wetlands, Oceans, and Watersheds, Oceans and Coastal Protection Division, USEPA, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1994b. *Methods for measuring the toxicity of sediment associated contaminants with estuarine and marine amphipods.* EPA-600/R-94/025, Narragansett, RI.

U.S. Environmental Protection Agency (USEPA). 1994c (draft). *Generic quality assurance project plan guidance for programs using community-level biological assessment in streams and rivers.* Prepared by Tetra Tech, Inc. for Assessment and Watershed Protection, USEPA, Washington, DC. August 1.

U.S. Environmental Protection Agency (USEPA). 1994d. *Guidance for the data quality objectives process.* EPA 600/R-96/055. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1994e. *Environmental Monitoring and Assessment Program: Integrated quality assurance project plan for the Surface Waters Resource Group, 1994*

activities, Rev. 200. U.S. Environmental Protection Agency, Las Vegas, NV. EPA 600-X-91-080.

U.S. Environmental Protection Agency (USEPA). 1993a. *Environmental Monitoring and Assessment Program. EMAP-Estuaries Louisianian Province. 1993 Quality Assurance Project Plan.* USEPA, Office of Research and Development, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1993b (draft final). *EPA requirements for quality assurance project plans for environmental data operations.* EPA QA/R-5. Quality Assurance Management Staff, USEPA, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1992. *Monitoring guidance for the national estuary program. Final.* EPA 842-B-92-004. Office of Water, Oceans, and Coastal Protection Division, USEPA, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1990. *Biological criteria: National program guidance for surface waters.* EPA-440/5-90-004. USEPA, Office of Water, Office of Regulations and Standards, Criteria and Standards Division, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1987. *Surface water monitoring: A framework for change.* USEPA, Office of Water, Office of Policy, Planning and Evaluation, Washington, DC.

U.S. Environmental Protection Agency (USEPA). 1986-1991. *Recommended protocols for measuring selected environmental variables in Puget Sound.* Looseleaf. Region 10, Puget Sound Estuary Program, USEPA, Seattle, WA.

Vernberg, F. S. 1972. Dissolved gases. Pages 1491-1526 *In* O. Kinne (editor). *Marine Ecology* , New York, Wiley-Interscience. Volume I, Part 3.

Ward, T.J. 1989. The accumulation and effects of metals in seagrass habitats. Pages 797-820 *In* A.W.D. Larkum, A.J. McComb, and S.A. Shepard (editors). *Biology of seagrasses. A treatise on the biology of seagrasses with special reference to the Australian region.* Elsevier, New York, NY.

Warwick, R.M. 1988. Analysis of community attributes of the macrobenthos of Fierfjord/Laugelundfjord at taxonomic levels higher than species. *Marine Ecology Progress Series* 46:167-170.

Warwick, R.M. and K.R. Clarke. 1991. A comparison of some methods for analyzing changes in benthic community structure. *Journal of the Marine Biological Association of the United Kingdom* 71:225-244.

Warwick, R.M., H.M. Platt, K.R. Clarke, J. Agard, and J. Gobin. 1990. Analysis of macroinvertebrate and macrobenthic community structures in relation to pollution and disturbance in Hamilton Harbor, Bermuda. *Journal of Experimental Marine Biology and Ecology* 138:119-142.

Washington, H.G. 1984. Diversity, biotic and similarity indices, a review with special relevance to aquatic ecosystems. *Water Research* 18: 653-694.

Watson, P.G. and T.E. Frickers. 1990. A multilevel, *in situ* pore-water sampler for use in intertidal sediments and laboratory microcosms. *Limnology and Oceanography* 35:1381-1389.

Weaver, G. 1984. PCB contamination in and around New Bedford, MA. *Estuaries* 18:22A-27A.

Weinstein, M.P., S.L. Weiss, and M.F. Walters. 1980. Multiple determinants of community structure in shallow marsh habitats, Cape Fear River Estuary, North Carolina, USA. *Marine Biology* 48:227-243.

Wiesberg, S.B., J.A. Ranasinghe, D.M. Dauer, L.C. Schaffner, R.J. Diaz, and J.B. Frithsen. 1997. An estuarine benthic Index of Biotic Integrity (B-IBI) for Chesapeake Bay. *Estuaries* 20:149-158.

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. 1993. *EMAP-Estuaries Virginian Province 1990 Demonstration Project Report.* EPA/620/R-93/006. Environmental Research Laboratory, USEPA, Narragansett, RI.

Wilding, J.R. 1940. A new square-foot aquatic sampler. *Limnological Society of America* Special Publication, No. 4:1-4.

Wonnacott, T.H. and R.J. Wonnacott. 1969. *Introductory Statistics. 2nd edition.* John Wiley and Sons, Inc., New York, NY.

Word, J. Q. 1980. *The infaunal trophic index.* The 1980 Annual Report, Southern California Coastal Research Project. Long Beach, CA. pp 19-39.

Word, J.Q. 1978. *The infaunal trophic index.* 1978 Annual Report, Southern California Coastal Water Research Project, Annual Report. Pages 19-39.

Word, J.Q., T.J. Kauwling, and A.J. Mearns. 1976. *A comparative field study of benthic sampling devices used in Southern California benthic surveys.* Southern California Coastal Water Research Project, 1500 East Imperial Highway, El Segundo, CA.

Wright, J.F., D. Moss, P.D. Armitage, and M.T. Furse. 1984. A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. *Freshwater Biology* 14:221-256.