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-cm^2 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).

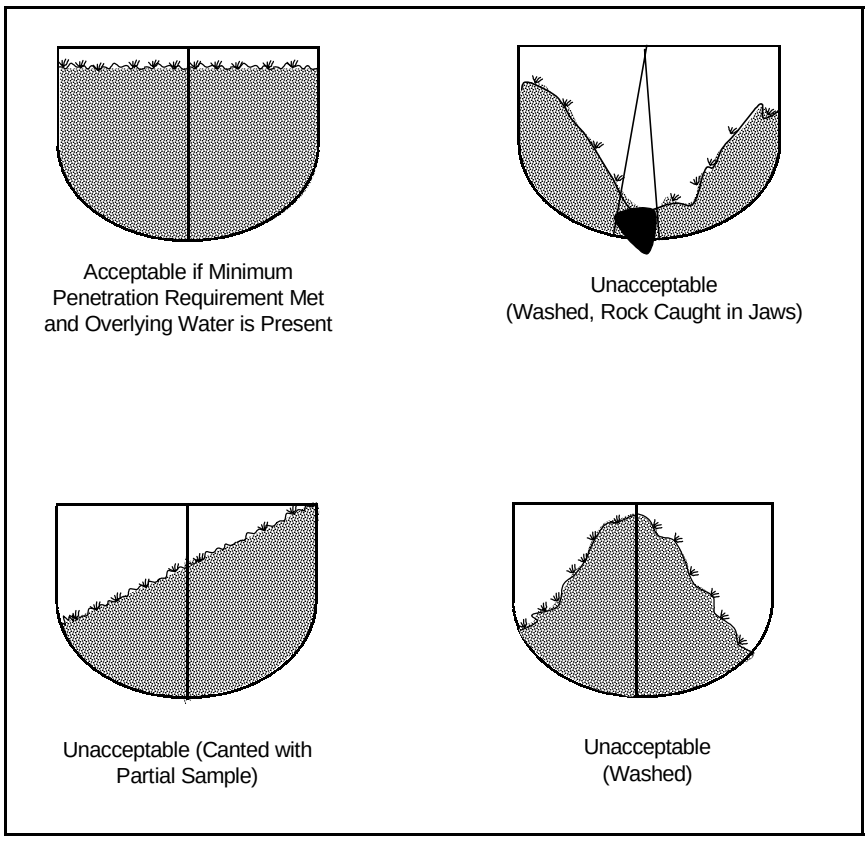
DEVICE	USE	ADVANTAGES	DISADVANTAGES
Ponar Grab Sampler	Useful on sand, silt, or clay.	Most universal grab sampler. Adequate on most substrates; very efficient for hard sediments. Large sample obtained intact permitting subsampling. Better penetration than other grabs; sideplates and screens reduce washout, 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.



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	substantially more or less than reference
dominant taxa	substantially more or less than reference
Tiers 2-3:	
% cover	reduced or enhanced
biomass	substantially more or less than reference
maximum depth of plant growth	reduced under enrichment
density of new shoots	reduced
stem counts	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).

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-β) 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

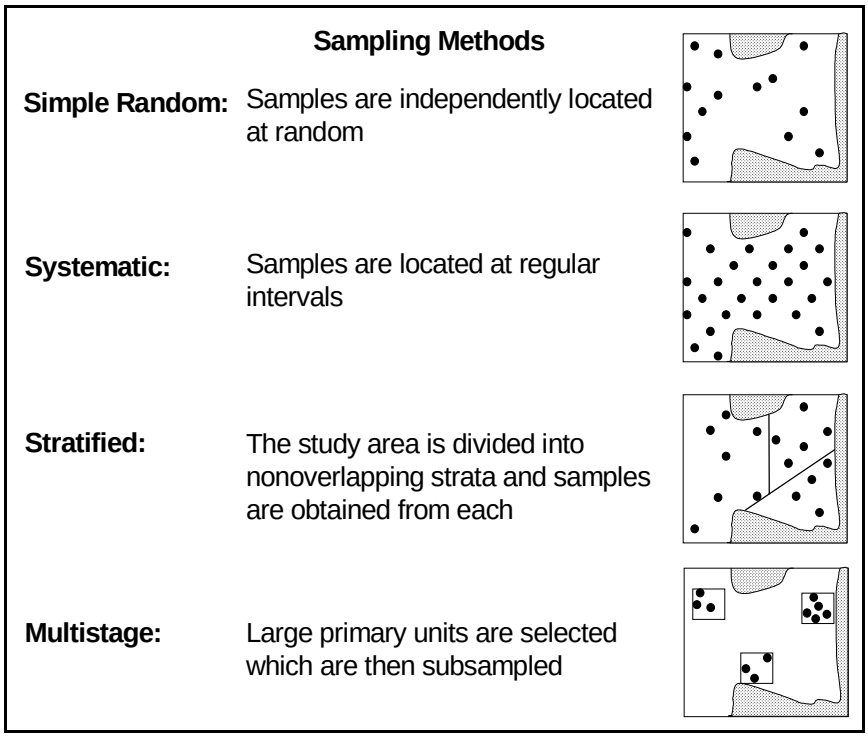


Figure 5-2
Description of various sampling methods. Adapted from USEPA 1992.

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 x 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.