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.

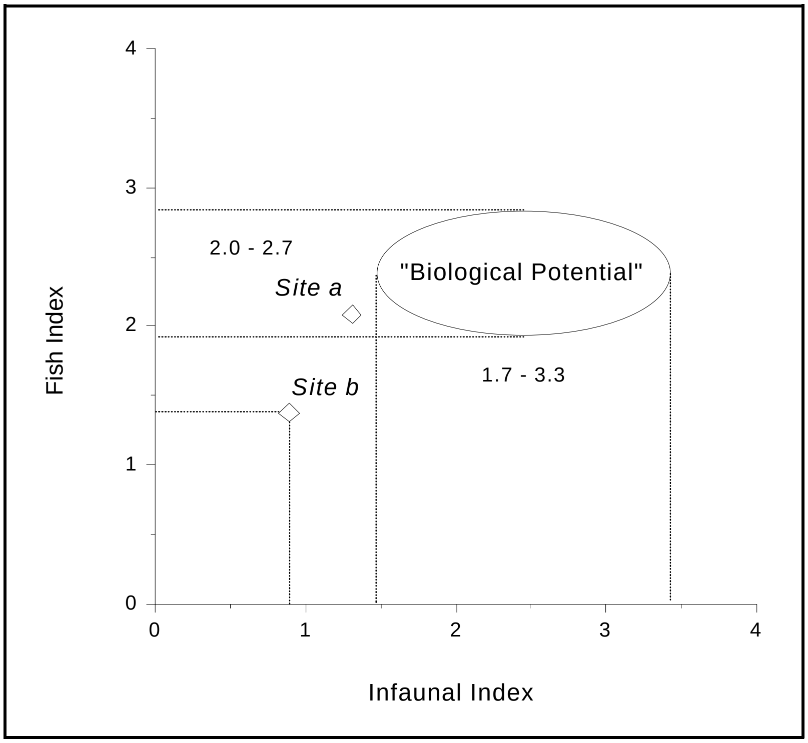


Figure 4-1
Graphical representation of bioassessment. Assessment sites *a* and *b* are compared to an ideal biological potential. Site *a* is near its potential. Site *b* deviates from it.

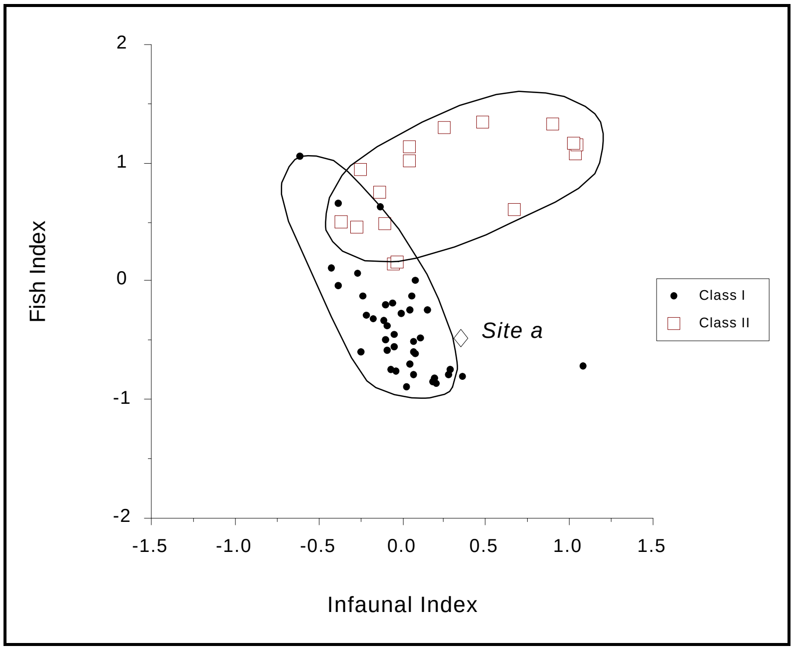
A population of reference sites might consist of sites which overlap different classes of estuarine or coastal marine waters (Figure 4-2). A useful classification system in this instance separates these reference sites into classes with different biological expectations. The classification itself must 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.

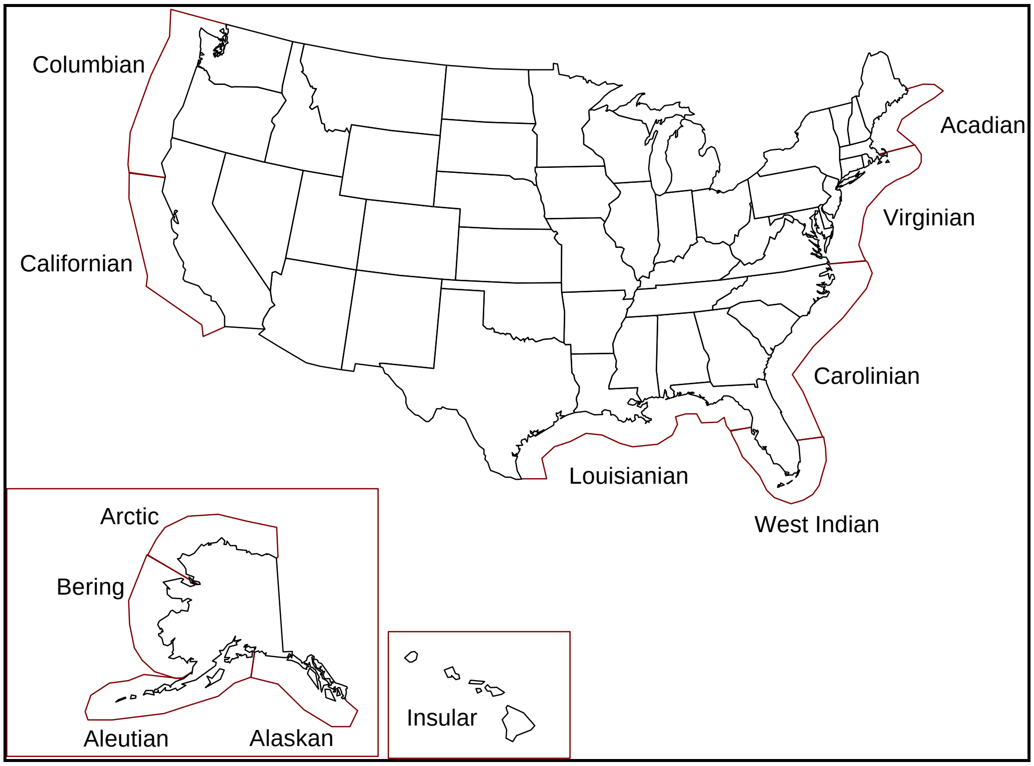
There are two types of coastal plain estuaries: classical and salt marsh. The classical coastal plain estuary is sometimes referred to as a "drowned river valley." These estuaries were formed during the last eustatic rise in sea level and they exhibit geomorphological features similar to river channels 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

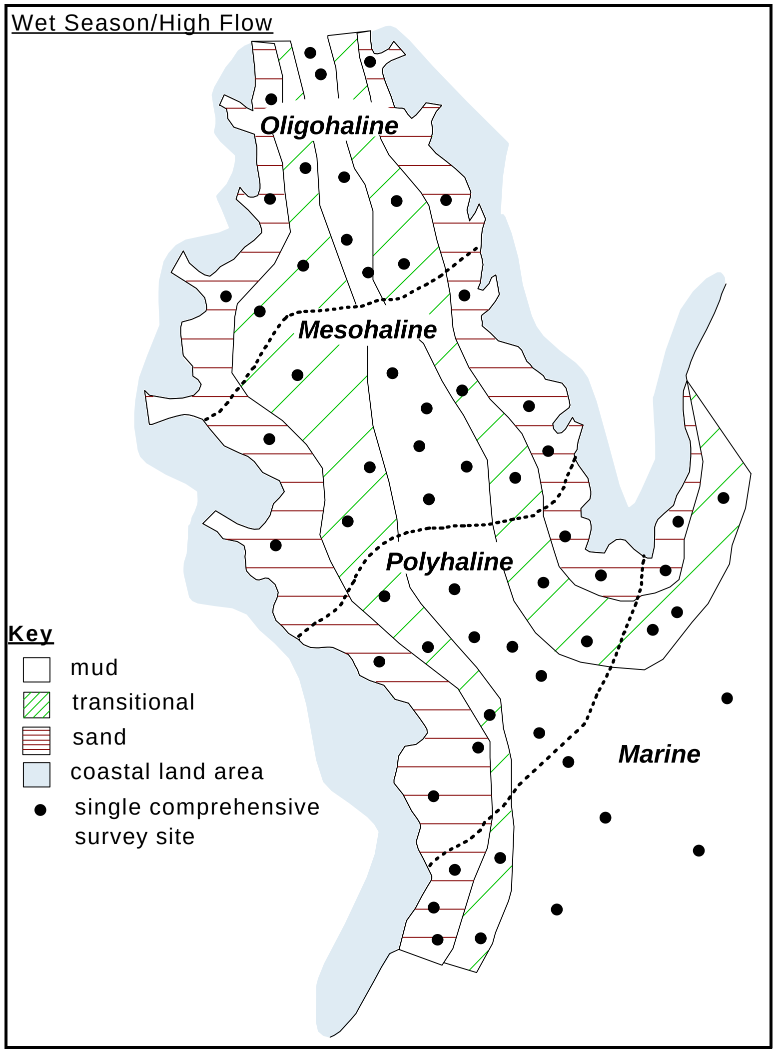
Estuarine and coastal marine reference conditions should be established using some combination of four elements: (1) evaluation of historical data; (2) sampling of reference sites; (3) prediction of expected conditions using models; and (4) expert consensus. Each element has its inherent strengths and 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.



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

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).

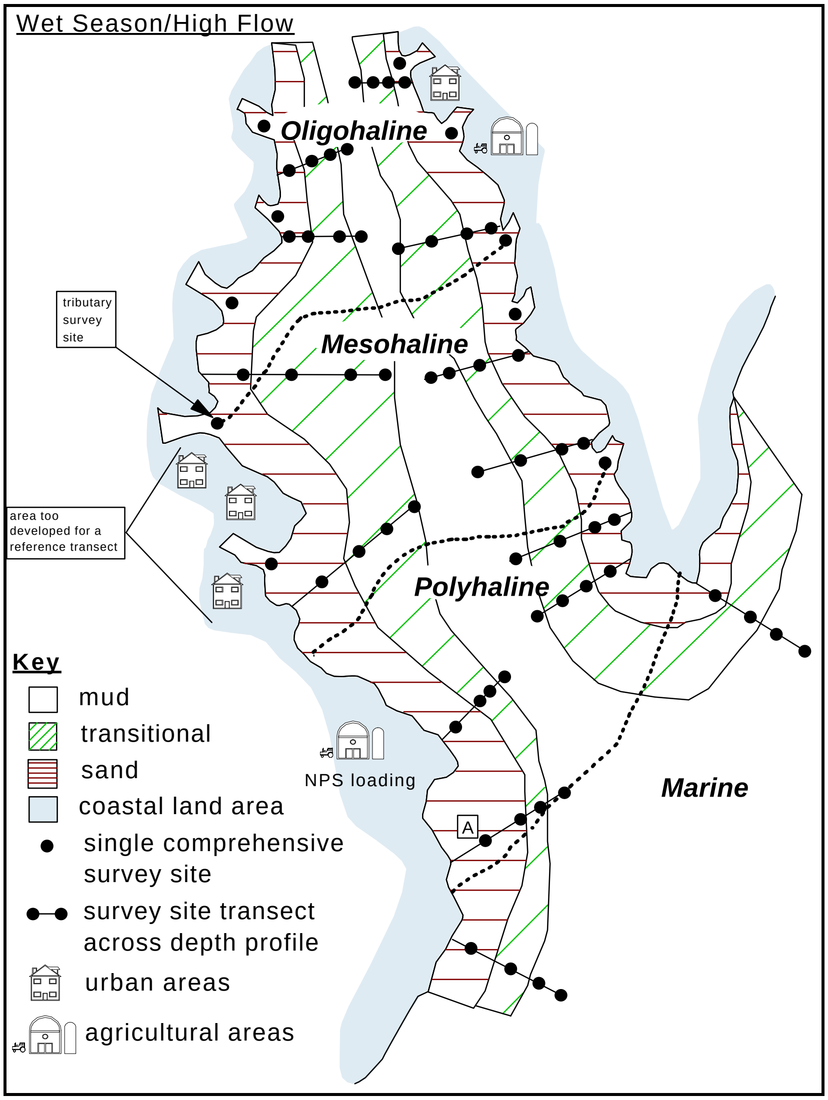
Stressed Sites - Effective metrics respond to environmental degradation and allow discrimination of impaired sites from the reference expectations. Metrics that do not respond are not useful in bioassessment. Response is determined by sampling a set of stressed sites in the same way as the reference 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.



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

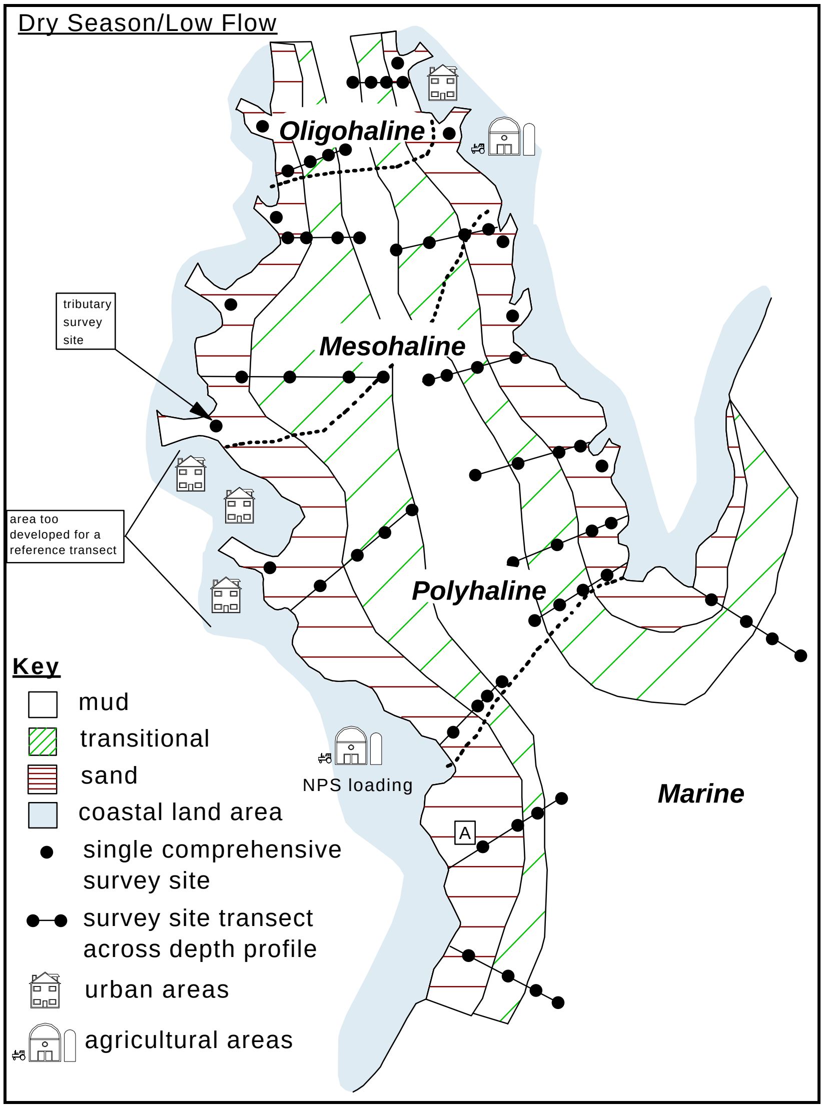


Figure 4-6

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

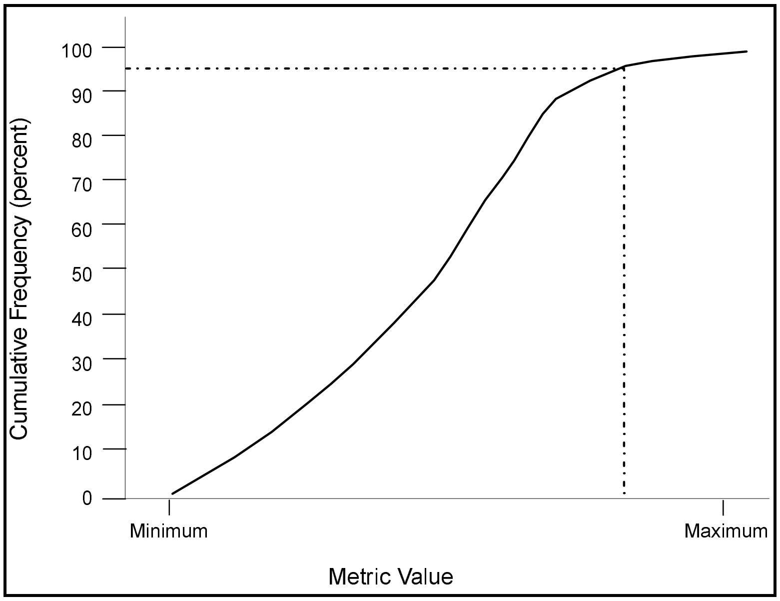


Figure 4-7
Hypothetical cumulative frequency distribution of metric values for all sites in a given estuarine or coastal marine class. The dotted line shows the metric value corresponding to the 95th percentile.

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.

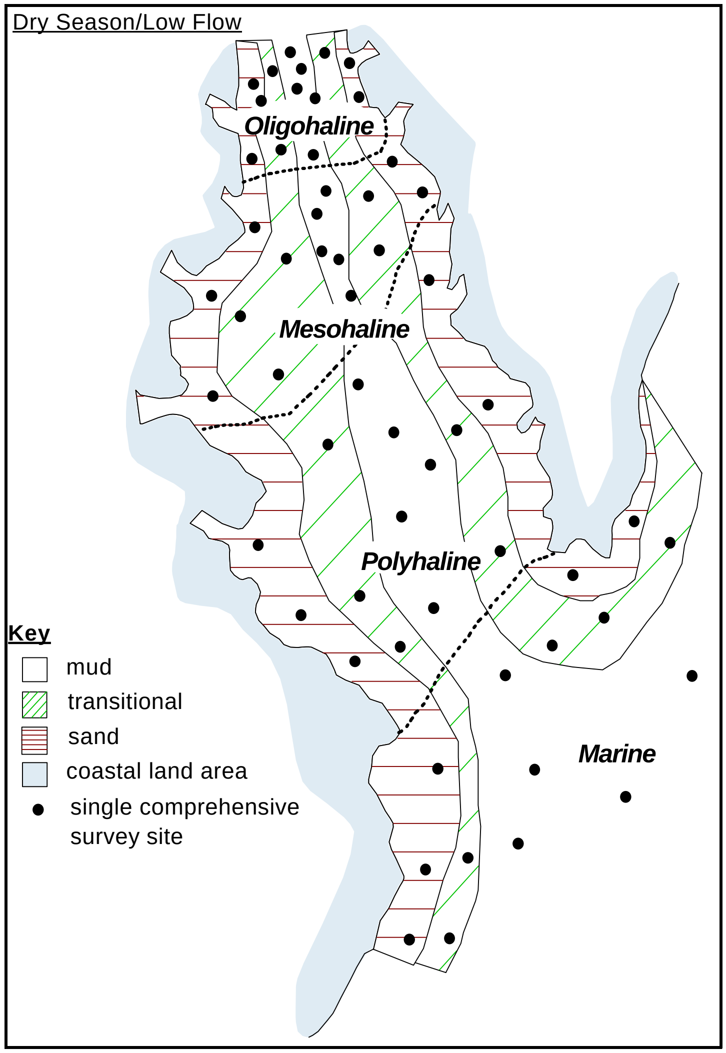
4.4.3 Site-specific Reference Sites

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.

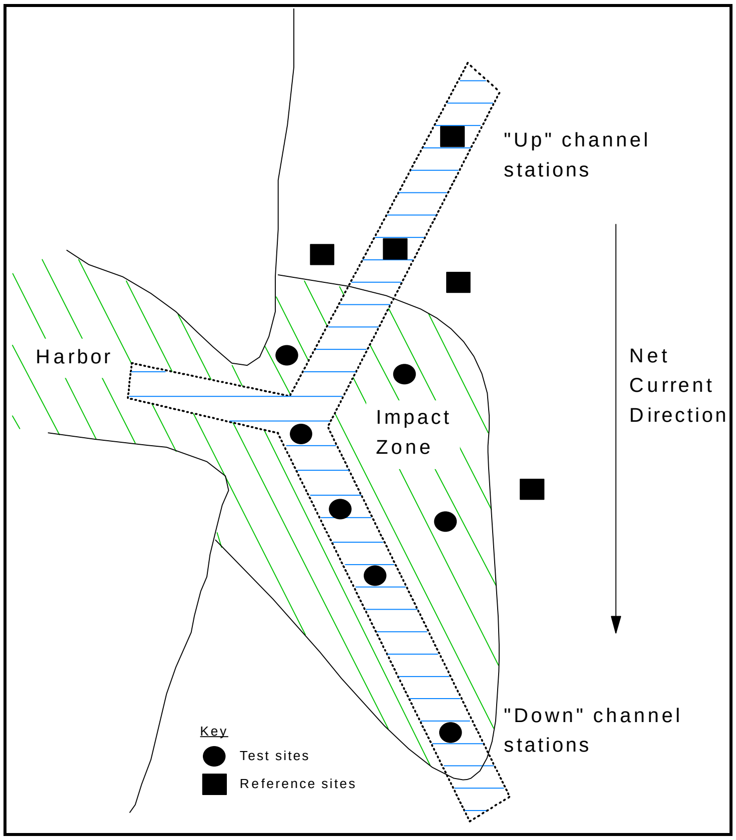


Figure 4-10

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