Please return to:

Dr. Bruce Wiersma
105 Winslow Hall
University of Maine
Orono, ME 04469-5782

MANAGING TROUBLED WATERS

THE ROLE OF MARINE ENVIRONMENTAL MONITORING

Committee on a Systems Assessment of
Marine Environmental Monitoring

Marine Board

Commission on Engineering and Technical Systems

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1990

In summary, the desired diagnostic national assessment of status and trends in marine environmental quality is most likely to come about through the orchestration, coordination, and synthesis of the results of well-designed local and regional studies. The inherent difficulties involved in comparing and accumulating the results of studies designed and conducted by different organizations and for different purposes are likely to be outweighed by the fact that studies tailored to specific environmental conditions and problems of an area have the best chance of yielding meaningful results. At the same time, although the committee can offer no panaceas or magic formulas, we urge continued efforts to achieve regional coordination of study protocols and parameters. Through development and demonstration of standardized approaches, such as those used in NOAA's NS&T Program, the federal government can encourage wider adoption of methods that will enhance the opportunities for development of information useful in national assessments.

4
Designing and Implementing Monitoring Programs

The technical design of monitoring programs refers to the process of deciding what to measure; how, where, and when to take the measurements; and how to analyze and interpret the resulting data. Proper analysis and interpretation of monitoring data result in information that helps scientists and managers decide whether regulatory, environmental quality, and human health objectives are being met. As emphasized in Chapter 2, when monitoring data have been converted to information in this manner, they generally provide better support for specific management actions. This chapter presents comprehensive guidance for developing the technical design of monitoring programs and describes a procedure for ensuring that the information produced meets the needs of managers and decision makers. This chapter is intended to guide those who implement monitoring programs toward better program design and improved dissemination of information gained from monitoring.

An appropriate technical design is critical to the success of monitoring programs because it provides the means for ensuring that data collection, analysis, and interpretation address management needs and objectives. To ensure that monitoring systems will produce information that is useful to decision makers, monitoring programs that address public concerns must be developed using a comprehensive methodology such as the one described here.

The committee emphasizes the importance of the following overall conclusion related to designing and implementing monitoring programs: *Failure to commit adequate resources of time, funding, and expertise to up-front*

program design and to the synthesis, interpretation, and reporting of information will result in failure of the entire program. Without this commitment, effort and money will be spent to collect data and produce information that may be useless.

A CONCEPTUAL APPROACH TO DESIGNING MONITORING PROGRAMS

Technical design can be challenging. Variability in nature creates "noise" that often obscures the "signal" of human-induced impacts. Multiple human activities occurring within the same area or time span can interact to create complex cumulative effects. Further, choices must be made among the wide array of scientific tools that could be used and the many environmental parameters that could be measured. For example, monitoring to measure degradation in fish communities could focus on the number of species in the community, community trophic structure, the incidence of abnormalities, or many other parameters.

The committee found no shortage of good advice concerning the technical design of monitoring programs. Such useful works as Holling (1978), Green (1979), Beanlands and Duinker (1983), Fritz, Rago, and Murarka (1980), NRC (1986), Wolfe (1988), Isom (1986), Rosenberg et al. (1981), Perry, Schaeffer, and Herricks (1987), and O'Connor and Flemer (1987) provide a rich resource of ideas, strategies, and technical methods. However, a major problem revealed in the case studies is a failure to apply the appropriate design tools consistently to fulfill clearly stated monitoring objectives. The case studies and the experience of committee members indicate that too little attention is directed at deciding what measurements are required to address the priority issues defined by the public and decision makers. Such priorities provide the context for selection and application of technical design strategies.

The comprehensive methodology presented here is drawn largely from the references cited above. The goal of this synthesis is to provide a methodology for formulating clear monitoring objectives at the outset; for designing statistically sound, cost-effective sampling programs consistent with those objectives; and for synthesizing, interpreting, and reporting monitoring data.

The following sections present a design methodology that is an expansion of the central elements of the conceptual framework shown in Figure 4.1. It provides a logical and scientifically based means of linking technical decisions about monitoring design to the information needs of the decision-making process. The methodology is generic and therefore applies to most monitoring situations.

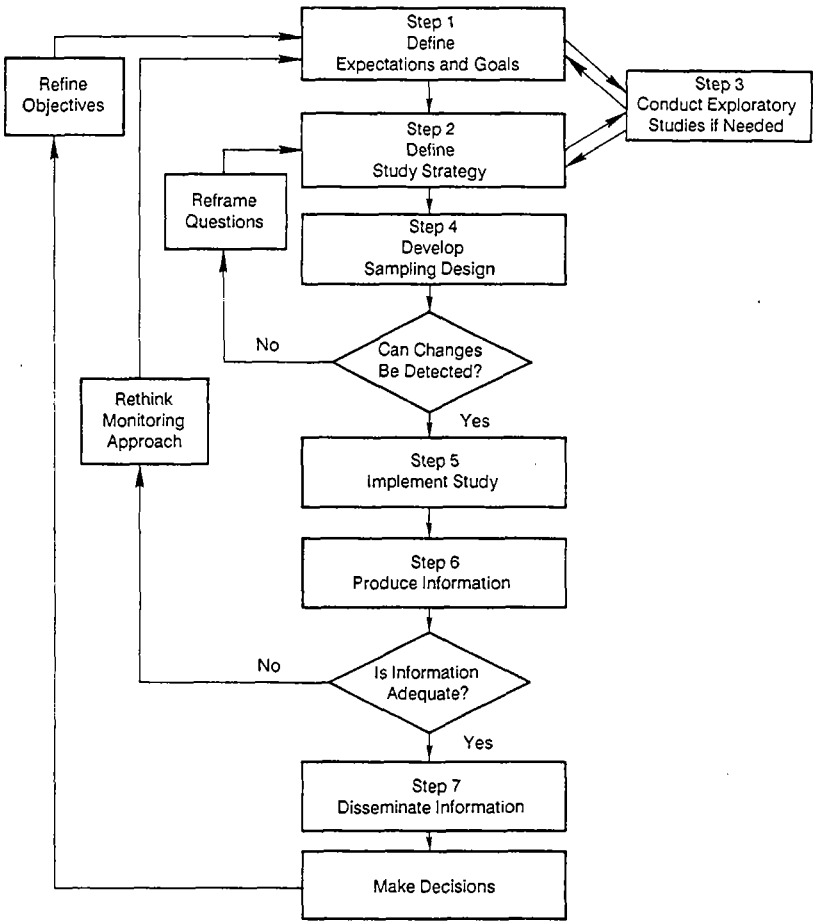

FIGURE 4.1 The elements of designing and implementing a monitoring program.

General Versus Specific Design Methodologies

A generic monitoring design methodology must be applicable to the various requirements of each monitoring category considered in this report—compliance, trends, and hypothesis testing. All three categories encompass a broad variety of questions about resources in many different habitats. In addition, resources, the processes that affect them, and human activities vary on diverse spatial and temporal scales. Too specific a methodology (i.e., one that specifies the exact models, parameters, sampling plans, and analyses) would be applicable only to a narrow range of

situations. Conversely, a methodology that is too general will not be useful to practitioners.

The committee resolved the conflict between the needs for specificity and for generality by developing a conceptual methodology that provides guidance in producing effective technical designs for most situations. The methodology does not furnish answers to all design problems. Instead, it identifies which problems are most important and describes how they can be solved. For example, it leads practitioners through steps that convert monitoring objectives into testable questions. It provides guidance in dealing with sources of variability and uncertainty and shows how feedback mechanisms help refine questions and objectives. It demonstrates methods for linking the collection and analysis of monitoring data to the information needs of the public and decision makers. Examples are used to demonstrate how elements of the methodology would be applied to specific situations. Some steps in the methodology are more relevant to some kinds of monitoring than others.

Despite its guidance, the methodology cannot replace local or specific scientific expertise. In fact, its successful application depends on the knowledge and skill of local experts. In this respect, it reflects the decision-making approach adopted by the U.S. Army Corps of Engineers (COE) for disposal of dredged material (Peddicord et al. in press; Cullinane et al. 1986).

A Methodology for Monitoring Design

Figure 4.1 shows the main elements of the conceptual methodology, each of which is discussed in detail in subsequent sections. The methodology is based on two principles: monitoring designs must reflect cause-effect relationships while accounting for variability and uncertainty, and specific design decisions (e.g., the number of stations and replicates to be collected) can be made only after objectives and related information needs are clearly established. A lack of clarity in purpose and expectations invariably results in failure to formulate a meaningful monitoring strategy (Green 1979).

Working upward from the bottom of Figure 4.1 helps in understanding the relationships among the steps in the methodology. Information can be disseminated to decision makers (step 7) only after it has been produced (step 6). Information is produced when the results of a carefully implemented study that includes adequate data analysis and interpretation have been summarized and evaluated (step 5). For a study to be implemented successfully (step 5), it must be designed (step 4) to develop answers to important questions effectively (step 2). The focused questions that serve as the basis of a monitoring program rely on clear management objectives

(step 1). Finally, preliminary studies (step 3) are often necessary to refine questions and technical aspects of the monitoring design.

Figure 4.1 also shows three important feedback points. The first, between steps 4 and 2, provides a means of reframing the study's underlying questions in light of real-world scientific, logistical, and financial constraints. As an example of such feedback, the Minerals Management Service (MMS) of the Department of the Interior evaluated historical data (Bernstein and Smith 1984) to help establish the objectives and design of a large-scale sampling program off California. The finding of this historical evaluation that natural variability made it extremely costly to detect changes in individual species helped focus the sampling program on other less variable and more sensitive parameters. The other feedback points in Figure 4.1 (encompassing steps 6, 7, and 1) allow program designers to review and modify monitoring objectives in light of actual monitoring information about the effectiveness of specific management actions and technological advances that occur during the study.

The above, and other, feedback points at more detailed levels of the methodology permit information that results from monitoring to be used to refine the sampling design. Throughout the more detailed description of the methodology that follows, feedback loops emphasize the point that information developed at one stage must be used to refine previous stages in an iterative process. For example, as scientific understanding and predictive ability increase, feedback mechanisms can be used for redirecting resources toward unanswered questions and away from issues that have already been addressed adequately. When such feedbacks are not used, monitoring loses its effectiveness for controlling and understanding human impacts on the environment. For example, electric utilities in Southern California continue to monitor for detrimental effects of thermal discharges from coastal power plants, even though nearly 20 years of monitoring have documented the limited consequences and spatial extent of thermal effects.

STEP 1: DEFINE EXPECTATIONS AND GOALS

As outlined in Chapter 2, the ultimate goal of monitoring is to produce information that is useful in making management decisions. Therefore two-way communication between scientists responsible for designing monitoring programs and the users of monitoring information is essential. These interactions give decision makers and managers an understanding of the limitations of monitoring and at the same time provide the technical experts who design monitoring programs with an understanding of what questions should be answered. Step 1 of the methodology (see Figure 4.2) is designed to ensure that this communication takes place in a structured context.

Such communication is important because anticipated population

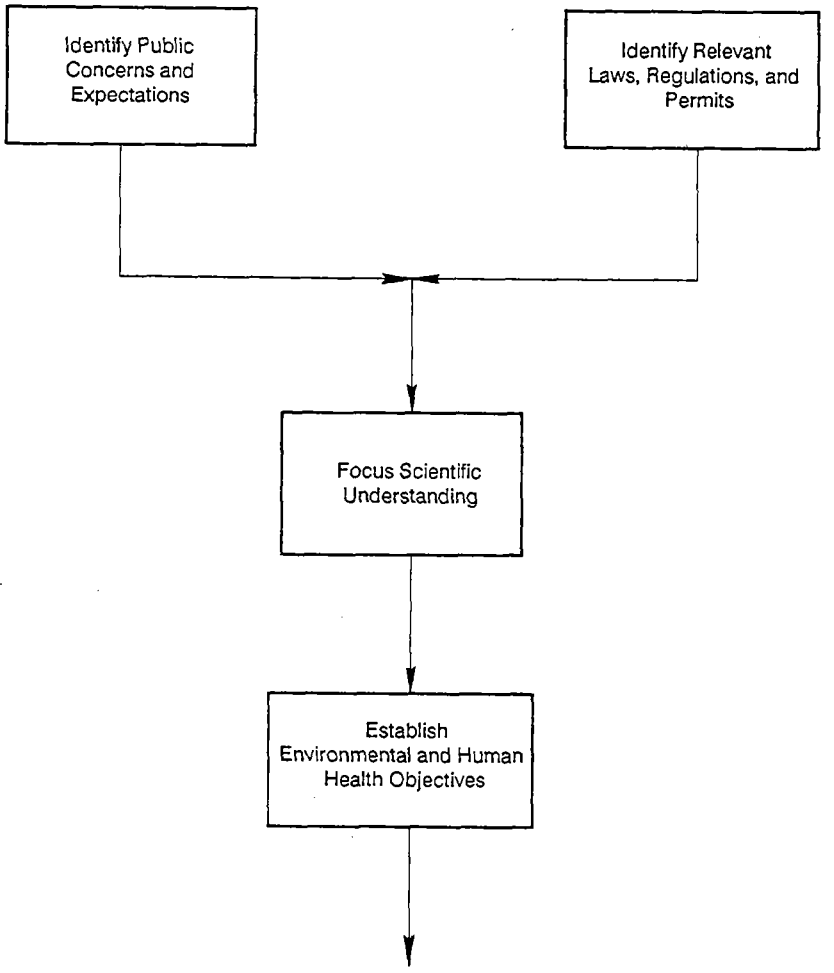

FIGURE 4.2 Step 1: define expectations and goals of monitoring.

growth and continued development of the coastal zone will increase the demand for monitoring information to support environmental decision making (EPA 1987; Champ, Conti, and Park 1989). If monitoring programs are to meet these demands, their objectives must integrate public concerns and expectations with the legal and regulatory framework through the use of scientific understanding to identify the relevant questions to be addressed.

BOX 4.1
A TECHNICAL DESIGN THAT MEETS MANAGEMENT NEEDS

DAMOS—the Disposal Area Monitoring System—collects only those data that can be shown, beforehand, to be useful in making management decisions or resolving technical problems (Fredette et al. in press). The DAMOS program clarifies and updates its definition of information needs through its technical advisory committee of independent scientists and through periodic public symposia. Although DAMOS has been criticized for not addressing larger-scale issues, such as the added stress of dredged material disposal on regional oxygen depletion, it has successfully addressed most important questions related to dredged material disposal. Most important, monitoring is fully integrated into the decision-making process, with active and ongoing interaction between those responsible for monitoring and those responsible for making decisions.

Just as the creation of useful information depends on clear monitoring objectives, these objectives depend on unambiguous statements about what constitutes useful management information (Cowell 1978). As Bernstein and Zalinski (1986) point out when talking about useful information, one must answer the questions "Information about what?" and "Useful to whom, and in what way, specifically?" Stating clear monitoring objectives involves answering these questions as precisely and unambiguously as possible.

The three case studies identified many instances in which the development of clear objectives helped translate monitoring data into information that supported management actions. An outstanding example is the DAMOS (Dredged Area Monitoring System) program carried out by the COE New England District to guide decisions about the disposal of dredged material (Fredette et al. in press; Engler and Mathis 1989). (See Box 4.1.)

In another instance, "tiered" monitoring (Fredette et al. in press; Zeller and Wastler 1986), exemplified by the monitoring plan for the 106-mile dumpsite off the East Coast (Werme et al. 1988), is structured to yield information that can answer a hierarchy of questions. Monitoring within the site concentrates on specific questions about the dispersal of disposed material. A finding that material has spread beyond the site boundary triggers a management action: more comprehensive monitoring to answer a higher tier of questions about environmental effects.

The Southern California Bight case study highlighted real-world impediments to developing clearly stated monitoring objectives. In the bight, multiple point and nonpoint sources of contaminants are in close proximity, and effects on a variety of important marine resources overlap. Marine resources in the bight are also affected by regionwide natural disturbances (e.g., El Niños, storms, and population blooms of organisms) that complicate the assessment of changes from human sources. It is much more difficult to document such cumulative effects than it is to measure those from single isolated sources or events. In addition, natural variation of resources and contaminants in the bight frequently occurs on spatial and temporal scales that confound the results of monitoring programs. The limited scientific understanding of how all these processes interact makes it difficult to find clear answers to many of the questions asked by decision makers and the public. All such impediments must be identified and considered when developing objectives for monitoring programs because they affect whether it is possible to fill the information needs identified in the definition of objectives.

Many approaches to defining issues and establishing monitoring objectives (see Figure 4.2) within the constraints imposed by the scientific knowledge base and resources (availability of time, money, and personnel) are possible (e.g., Adamus and Clough 1978; Capuzzo and Kester 1987; Gilliland and Risser 1977; Walker and Norton 1982; Wiersma et al. 1984; Cairns, Dickson, and Maki 1978)). Results of one approach (Clark 1986) that was found by the Southern California Bight case study to be especially useful are summarized in Figure 4.3. This cumulative assessment approach presents a synoptic picture of natural and human sources of disturbance and impacts and their effects on natural resources. Conducting this kind of analysis requires making decisions about which resources are valued and/or vulnerable. It also requires synthesizing available scientific information about how they are impacted. A particularly useful aspect of this approach is the identification of multiple and cumulative impacts. Further, it includes information about the limits of scientific certainty associated with potential impacts. This procedure provided a framework for synthesizing available scientific information on the Southern California Bight in a way that could be used by scientists, environmental decision makers, and the public to begin establishing realistic monitoring objectives.

Even though the analysis underlying Figure 4.3 was qualitative and was based on incomplete understanding, it helped participants in the Southern California Bight case study identify potential effects not addressed by ongoing monitoring programs. Figure 4.3 was especially valuable as a tool for synthesizing the available information into a conceptual model of system interactions. This model thus provides an effective starting point

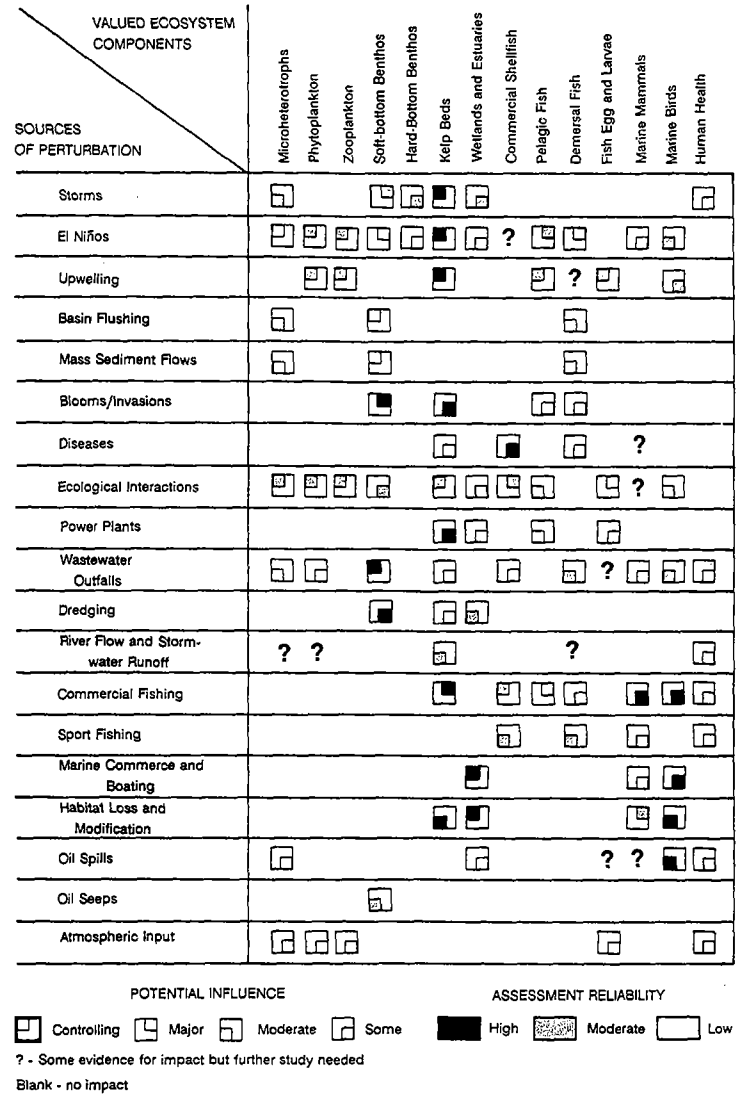

FIGURE 4.3 Impacts on the marine environment of the Southern California Bight. Note: Individual matrix cells illustrate the presumed relative impact of each source on each component, along with the associated scientific certainty. Columns represent cumulative impacts on individual components; rows represent the effects of individual perturbations on all components. This figure was used to summarize and investigate ways of identifying and ranking impacts in the Southern California Bight. SOURCE: Adapted from Clark 1986.

for developing monitoring objectives, including the selection of specific resources, impacts, and changes that should be monitored.

STEP 2: DEFINE STUDY STRATEGY

Figure 4.4 shows the elements of defining a monitoring strategy and developing specific questions to be answered. These questions guide subsequent steps in the technical design process. Step 2 begins with the general monitoring objectives developed in step 1 and ends with explicit questions to be answered that are the basis for developing a sampling design. The goal of this step is to narrow the focus of monitoring from the vast number of questions and parameters that could be examined to those that will produce the specific information needed. Step 2 is essential because, without clearly stated testable questions, monitoring is often a haphazard collection of data. As Green (1979) emphasizes, "Your results will be as coherent and as comprehensible as your initial conception of the problem." Similarly, in writing about monitoring to detect power plant impacts, Fritz, Rago, and Murarka (1980) stated: "This failure [to formulate clear-cut questions] may account for the relatively inconclusive results produced in environmental assessments."

There are no simple guidelines for producing specific questions to be answered. Whatever method is used, it must be pursued with the determination to continue until specific potential impacts on specific resources in specific locations at specific times are identified (e.g., Bain et al. 1986). To be useful, testable questions need not be complex; DAMOS managers were concerned about whether hurricanes would erode dredged material disposal mounds and contribute to the transport and dispersal of contaminants contained in the dredged material (SAIC 1986). Their concern led to the question "Within the detection limits of seabed profiling technology, are disposal mounds in Long Island Sound smaller after a hurricane than they were before the hurricane?" In contrast, the monitoring conducted around oil platforms in the Gulf of Mexico was not based on specific questions designed to meet specific information needs, lacked any operational definition of impact, implicitly assumed that impacts would be easily distinguishable from natural variation, and failed to use an appropriate sampling design. (See Box 4.2.)

In their study of impact assessment methods, Beanlands and Duinker (1983) provide a particularly good example of the difference between useful and nebulous questions. The original nebulous question "What would be the impacts of a proposed dam on the fish resources of the river?" failed to help focus the sampling design because it did not ask "What impacts and which fish resources are of concern?" Beanlands and Duinker explain how this original question was refined to provide the specific information

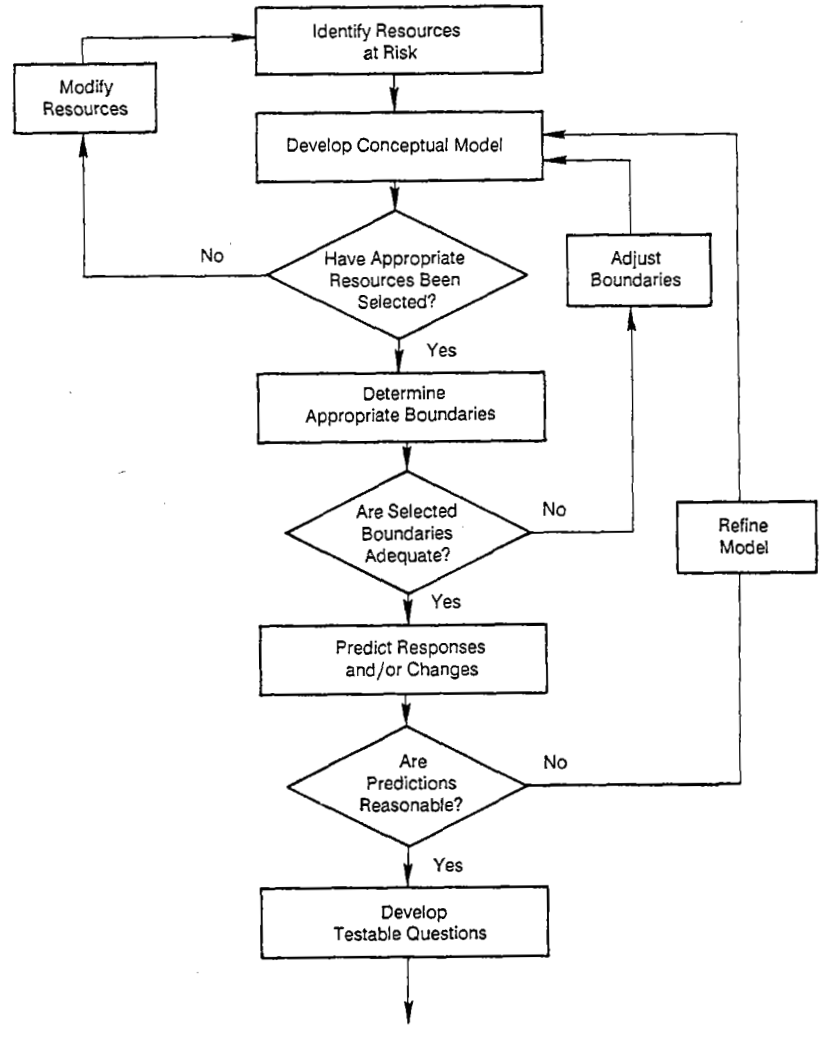

FIGURE 4.4 Step 2: Define study strategy.

needed to make a decision. The refined question was: "What percentage of the Arctic char spawning habitat would be lost given a 0.5 meter reduction in the water level of the river during the month of September?"

As shown in Figure 4.4, several steps are involved in progressing from general monitoring objectives (step 1 and Figure 4.3) to specific questions to be answered (step 2 and Figure 4.4). They include: identifying

BOX 4.2
THE EFFECTS OF OFFSHORE PETROLEUM ACTIVITIES

The environmental effects of offshore oil and gas exploration and production activities have engendered much concern since national policy promoted expansion of Outer Continental Shelf (OCS) development as a response to the Arab oil embargo and threats to U.S. energy security. These concerns focused not only on oil spills but on the effects of operational discharges—drilling fluids and cuttings and produced waters.

Few long-term monitoring programs have dealt with the effects of OCS development activities; several field assessments, although not continued for long periods, share a basic purpose and many design considerations with environmental monitoring. They include several studies in the Gulf of Mexico, which has experienced offshore oil and gas development since the 1940s, and monitoring studies in the Mid-Atlantic Bight, on Georges Bank, and, more recently, in central California.

Carney (1987) reviewed the design of these studies, conducted from 1972 to 1984 and sponsored by industry (off Louisiana and in the Mid-Atlantic Bight), the Environmental Protection Agency and National Oceanic and Atmospheric Administration (off Texas), and the Minerals Management Service of the Department of the Interior (off Louisiana and on Georges Bank).

Carney concluded that the three Gulf studies could not detect long-term impacts because they lacked an operational definition of impact, implicitly assumed that any impact would be easily distinguishable from natural variation, and failed to use design techniques afforded by population survey statistics. He characterized them as the "survey and explain" approach. Although the Mid-Atlantic and Georges Bank studies showed more thoughtful design, the Mid-Atlantic study had to abandon its statistical design because the model was ecologically unrealistic.

Carney concluded that OCS monitoring could be improved by concisely stating the problem, carrying out preliminary sampling, verifying the appropriateness of the sampling unit and estimating replication needed to obtain required precision, selecting and adhering to the adopted design and the results obtained, adopting a stratified random approach in the presence of large-scale environmental variation, taking randomly allocated replicates within each combination of controlled variables, and using replication to estimate variability.

the environmental components or resources at risk, establishing the links (direct and indirect) among ecosystem attributes, particularly the resources at risk and human and natural causes of change, establishing boundaries for spatial, temporal, biological, and physical/chemical aspects of the system (including defining scales of spatial and temporal variations in both human and natural causes of change); and projecting, either quantitatively or qualitatively, human and natural changes in resources and the interactions between them. These steps help define the cause-effect relationships (i.e., a quantitative or qualitative conceptual model) that determine potential responses of the resources at risk to human activities.

An essential part of developing a sampling design and study strategy is ensuring that sufficient feedbacks are incorporated so that the questions to be answered are refined to reflect the best information available, including new information that results when a study is implemented. The evolution of the sampling design of the San Onofre kelp bed monitoring program (see Box 4.3) is a good example of a typical situation in which information acquired during the technical design stage modified the original understanding of the system and thus the monitoring objectives. The example

BOX 4.3
SAN ONOFRE KELP BED MONITORING

How are testable questions developed, and how do real-world constraints affect technical design? An example should help.

The Southern California Edison Co. (SCE) San Onofre Nuclear Generating Station (SONGS) discharges cooling water (6,274 cubic meters per second) from offshore diffusers. The diffuser flow entrains 10 times the initial volume to create a plume that extends offshore, upcoast, or downcoast depending on local currents. SCE and the California Coastal Commission's Marine Review Committee (MRC) carried out extensive predictive and monitoring studies.

Preliminary studies (tank tests, numerical modeling, and field verification) showed that, at times, the plume would enter the nearby San Onofre kelp bed (*identify resources at risk*).* Further plume modeling and preliminary experiments on kelp reproductive processes (*preliminary studies*) showed that the plume's turbidity could affect light levels in the kelp bed and reduce reproductive success and that this change might reduce the size and long-term viability of the kelp bed.

These preliminary studies led to development of the *conceptual model.* It defined a causal link between power plant operation and kelp bed change, but it did not specify the extent and frequency of

impacts (*spatial and temporal boundaries*). *Making predictions* about the extent and frequency of the potential impacts required integrating additional ecological information about kelp reproduction with further plume model results. Natural year-to-year variability in kelp recruitment was large, suggesting that monitoring should continue over several years so that natural factors controlling reproductive success could be identified and quantified. Plume models indicated that only the portion of the kelp bed nearest the diffusers would be affected. In addition, reduced light levels due to the plume were estimated.

Predictions from the *conceptual model* and results from *preliminary studies* and other research were integrated to produce the following *specific questions* to be answered:

- Would the diffusers create a turbid plume that would enter the San Onofre kelp bed?
- Would a turbid plume reduce near-bottom irradiance?
- Would reduced near-bottom irradiance reduce successful kelp recruitment compared to areas not influenced by the plume?
- Would a succession of poor recruitment events reduce the size of the kelp bed?
- Would a smaller kelp bed be less viable over the long term?

Field studies validated numerical models of plume behavior and kelp reproduction. Compliance monitoring of the thermal plume documented plume behavior. Trends monitoring of kelp recruitment and kelp bed size tested hypotheses about both short- and long-term plume effects.

Several important points about step 2 of the technical design framework are shown here. The *specific questions* to be answered were more precise than merely asking "Does the SONGS power plant affect the San Onofre kelp bed?" To produce precise questions, SCE and MRC integrated preliminary studies with research and monitoring information on kelp bed ecology. Arriving at *specific questions* to be answered required refinement of the conceptual model and several iterations of modeling, research, and prediction. The first version of the conceptual model was verbal and qualitative and guided subsequent design efforts. Later versions incorporated ecological knowledge, quantitative predictions from numerical models, and subjective judgments.

*Italicized phrases indicate specific framework elements (Figure 4.4).

also demonstrates how preliminary studies can provide information crucial to developing and refining the specific questions to be answered.

Conceptual Models and Predictions

A description (i.e., a conceptual model) of the cause-effect links between human activity and anticipated environmental change is the central feature in developing specific questions to be answered. It is the conceptual model that is the means of predicting environmental change and the results of management action—predictions that efficiently direct and focus monitoring efforts. (See Box 4.4.)

A conceptual model describes links among the resources at risk; the physical, chemical, and biological attributes of the ecosystem; and human and natural causes of change. The understanding that results permits testable questions to be clearly stated and ultimately evaluated. By providing a context for organizing existing scientific understanding, a conceptual model also identifies important sources of uncertainty. As the San Onofre kelp bed example demonstrates (see Box 4.3), conceptual models can be qualitative or quantitative, depending on the state of knowledge, and they should be refined during a monitoring study.

Boundaries

Many workers (Cooper and Zedler 1980; De Angelis 1980; Dooley 1979; Fritz, Rago, and Murarka 1980; Hilborn, Holling, and Walters 1980; Beanlands and Duinker 1983; Green 1979; Holling 1978) emphasize the importance of establishing boundaries in monitoring and environmental assessment studies. These boundaries affect the kinds of questions that monitoring can answer. Suitable boundaries ensure that monitoring is relevant to both natural processes and the environmental quality and human health objectives established early in the technical design. For example, a major finding of the Southern California Bight case study was that the assortment of individual monitoring programs in Southern California was not adequate to address important potential regional and cumulative environmental effects. The Particulate Waste case study found that the site-specific monitoring programs in Long Island Sound could not easily relate changes at individual sites to regional hypoxic events. In both instances, the boundaries of the monitoring programs reflected site-specific rather than larger-scale questions.

Adequate boundaries make it more likely that all events and processes seriously affecting the system will be included. For example, water quality in the upper Chesapeake Bay is affected by the volume of freshwater input from rivers (Holland, Shaughnessy, and Hiegel 1986). Freshwater input,

BOX 4.4
CONCEPTUAL MODELS

The term "conceptual model" is sometimes misunderstood. In the technical design methodology, conceptual models refer to descriptions of causes and effects that define how environmental changes are expected to occur. Conceptual models may incorporate disparate elements (e.g., numerical models, natural history information, ecological theory, and subjective judgment) (Holling 1978; Goodall 1977; Pielou 1981; Saila 1979). Deciding just what tools to use depends on monitoring objectives, available technical knowledge, and the degree of precision required.

A wide variety of conceptual models is currently used to structure marine monitoring practice because the best representation of a specific system depends on the nature of the perturbation and the resources at risk (Beanlands and Duinker 1983; NRC 1986). The examples below show some of the approaches.

Monitoring in the San Onofre kelp bed (see Box 4.3) was based on a conceptual model that included kelp physiology and physical transport processes. When entrainment of fish eggs and larvae and impingement of adult fish by coastal power plants is a concern, conceptual models are based on population dynamics and circulation patterns that control larval transport (Barnthouse, DeAngelis, and Christensen 1979; Barnthouse and Van Winkle 1981; Boreman, Goodyear, and Christensen 1981; Polgar, Summers, and Haire 1980; Polgar, Turner, and Summers 1988).

Monitoring toxic effects of discharged chemicals is typically based on a source-receptor conceptual approach (e.g., Wiersma and Otis 1986; Wiersma et al. 1984; Behar et al. 1979; Eberhardt et al. 1976; Pickett and Whiting 1981). The ultimate receptor in such a source-receptor model could be a physical ecosystem compartment, a particular species, or an organ (e.g., the liver) within a species.

In all these approaches, the conceptual model began as a qualitative description of causal links in the system. Based on available technical knowledge, it was then expanded to include quantitative elements, such as analytical or numerical models. Sometimes the conceptual model includes more than one kind of numerical model (e.g., in the kelp bed example, plume dispersion and light attenuation models).

in turn, varies with major storms and with seasonal patterns of rainfall in the watersheds around the bay. Given this information, it is clear that the monitoring program boundaries needed to include surrounding watersheds.

Setting boundaries is not typically separate of the monitoring program's other elements. Boundaries are implicitly set when environmental quality and human health objectives are established and resources at risk identified. Additional boundaries are set when a conceptual model of linkages among ecosystem components is constructed. Walters (1986) points out that the defined boundaries of the system and the specific questions that can be asked are interdependent. The initial questions help define the system boundaries, but changing the boundaries also affects the questions that can be asked. Thus Walters suggests varying boundaries deliberately during planning, offering a valuable opportunity to examine a given situation from several perspectives. Even though boundary setting is part of several steps in the methodology, it is so important to the overall effectiveness of monitoring that explicit evaluation of boundaries is shown as a separate activity. (See Figure 4.4.) This step ensures that boundaries will result from conscious decisions rather than from unstated assumptions.

Several approaches to establishing study boundaries have been proposed; three of them, which are described below, illustrate the factors that should be considered. No one approach or one set of boundaries is best for all problems.

Beanlands and Duinker (1983) identified four types of boundaries that contribute to the sampling design of a monitoring program: (1) administrative boundaries deriving from political, social, or economic factors; (2) project boundaries deriving from the spatial and temporal extent of the project or perturbation under study; (3) ecological boundaries deriving from the nature of physical, chemical, and biological processes; and (4) technical boundaries deriving from limits on capabilities to predict or measure ecological change and/or ecosystem processes. The final study boundaries that are established should be consistent with these classes of boundaries and with study objectives. Walters (1986) discusses four dimensions to consider when establishing boundaries: the breadth of factors considered, the depth of analysis, the spatial scale of variables, and the time scale or horizon. Figure 4.5 shows another, visual, method for identifying the relative space and time scales of important ecosystem components and processes that are relevant for establishing study boundaries.

Establishing appropriate space and time boundaries is particularly important in developing the sampling design for monitoring studies for several reasons. First, the majority of parameters that could be measured by monitoring programs vary on space and time scales. No one set of boundaries is adequate for all parameters. Second, events that occur over large areas typically occur over long periods and vice versa (e.g., Chelton, Bernal, and

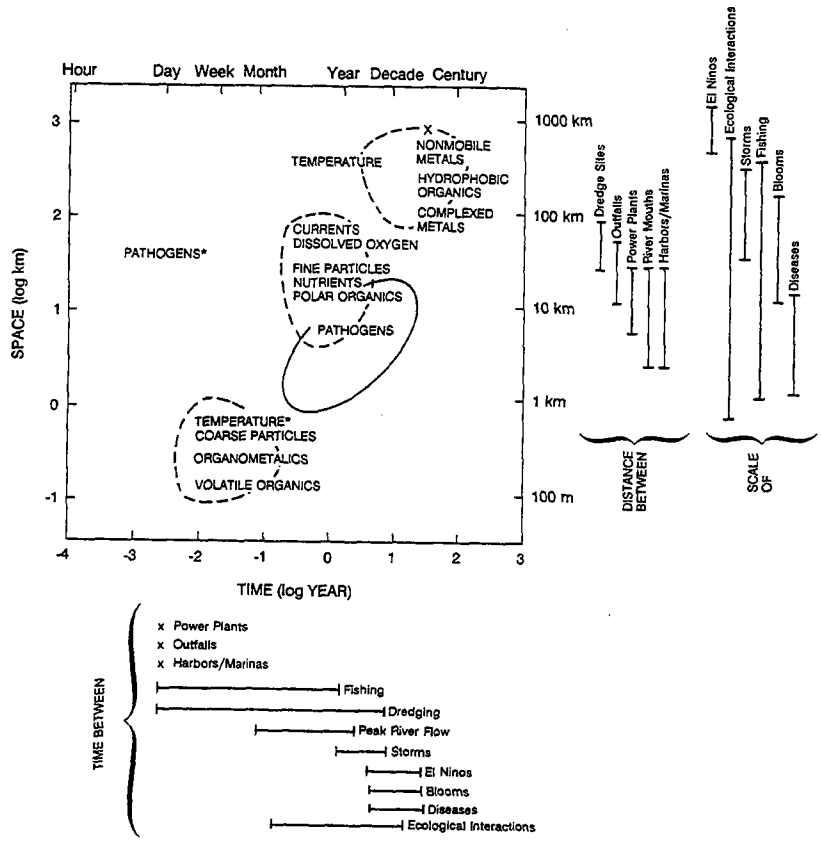

FIGURE 4.5 Characteristic temporal and spatial scales of the movement of marine constituents and the processes of sources of perturbation. SOURCE: Adapted from Clark 1986.

McGowan 1982; Haury, McGowan, and Wiebe 1978). And third, spatial and temporal variability in the marine environment can easily confound the interpretation of monitoring results (Botkin and Sobel 1976; Livingston 1982; Pearson and Barnett 1987; Holland, Shaughnessy, and Hiegel 1986). Establishment of appropriate boundaries ensures consideration of all events and processes that affect the questions being asked and thus the sampling design.

Predictions and Uncertainty

It is important to recognize that the sampling designs of all monitoring programs are based on assumptions and predictions about likely responses to perturbation. In addition, management actions are taken with the expectation that they will mitigate an impact or protect a resource. Ideally, all such predictions should be made explicitly. For example, decisions about waste load allocations in the New York Bight were based on numerical models that predicted the effects of alternative management actions (Hazen and Sawyer and Hydroscience 1978). More often, predictions are not made explicitly, but even then, the assumptions implicit in the choice of one alternative over others constrain monitoring and management actions just as though they had been made explicitly (Beanlands and Duinker 1983; Bernstein and Zalinski 1986; Holling 1978). For example, monitoring once a year precludes monitoring seasonal patterns, just as though it was assumed that seasonal patterns do not significantly affect year-to-year trends. Similarly, monitoring only benthic infauna precludes identifying impacts on other fauna; the implicit assumption is that benthic infauna are the best indicators of impacts.

In general, two types of predictions are important: those about how a particular environmental perturbation will affect a monitored parameter and those about how a specific management action will affect an important resource. Predictions should be based on a conceptual mode that specifies cause-effect relationships, and they should be stated clearly. Implicit assumptions should be avoided because their consequences are frequently not evident either to the designers or to the users of monitoring data.

Numerical models are often used to make predictions for monitoring design and management because they systemize knowledge and produce quantitative predictions. However, it is important to recognize that numerical models are not infallible. When they are wrong, the mismatch between prediction and experience affords an opportunity to improve understanding.

As one example, water quality models were used to predict the effects of proposed freshwater diversions on productivity in San Francisco Bay (e.g., Hydroscience 1974). The drought of 1976-1977 tested these predictions; it mimicked the reduced flow due to diversions. The altered salinity regime affected benthic communities, resulting in a new assemblage. This new assemblage cropped phytoplankton at an increased rate, resulting in much lower plankton productivity than predicted (Nichols 1985b).

As another example, the 1976 impoundment of a Manitoba lake (Lehman 1986) revealed serious flaws in models of reservoir dynamics. Unforeseen physical effects led to extreme biological changes, and higher water levels melted permafrost and led to severe erosion. Increased turbidity affected productivity in unexpected ways, and the commercial whitefish

fishery collapsed. Mercury released during erosion was mobilized biologically, and tissue levels in walleye and northern pike rose above Canadian standards.

Uncertainty associated with assumptions and predictions varies in kind and degree (Holling 1973, 1978; NRC 1986; Walters 1986; Wolfe 1988; Freudenburg 1988; Eberhardt 1976; Paine 1981). Indeed, were there no uncertainty about environmental effects, monitoring would be unnecessary. Identifying the sources and consequences of uncertainty is sometimes more valuable than predicting the impacts (NRC 1986). Uncertainty stems from many sources (e.g., model error, measurement error, natural variability, incomplete scientific understanding) and takes many forms. At one extreme, uncertainty in well-understood systems is mainly a result of measurement error. At the other extreme, it may be difficult or impossible to predict whether an assumed effect will occur at all. In many cases, the nature of effects is clear, and only their timing, extent, or severity is uncertain. For example, the effects of organic enrichment on benthic communities are well documented (e.g., Pearson and Rosenberg 1978; Greene and Sarason 1974; Greene and Smith 1975; Stull et al. 1986; Pearson 1987; Norton and Champ 1989; Swartz et al. 1986). But the actual severity and extent of changes that will occur in any specific instance are not predictable.

The lack of clear predictions is probably the single greatest weakness of most monitoring programs. In spite of its difficulty, basing the monitoring design on explicit predictions results in discernible benefits. It provides an unambiguous statement of the assumptions underlying monitoring and management actions and increases the likelihood that all important processes and interactions will be considered adequately. It ensures that monitoring will be focused on expected changes that are relevant to management objectives. Further, explicit predictions of the kinds and amounts of change expected provide the basis for a statistically rigorous sampling, measurement, and analysis design.

STEP 3: CONDUCT PRELIMINARY STUDIES AND RESEARCH

Figure 4.1 shows that preliminary studies (step 3) support the development of both specific questions to be addressed (step 2) and the sampling designs (step 4). Such studies help refine measurement techniques, result in the development of new methods or models, estimate the magnitude of natural variability, or otherwise lay the groundwork for developing a monitoring design.

Preliminary studies have supported development of the technical design of many monitoring programs. In the kelp bed example (see Box 4.2), preliminary studies included numerical and physical modeling of the diffuser plume, physiological studies of kelp reproduction and recruitment, and

field measurements of natural kelp distribution, growth, and mortality. In Chesapeake Bay, research to characterize and understand natural variability in benthos and field experiments to validate their responses to predation by fish and crabs and the estuarine salt gradient (Holland 1985; Holland, Mountford, and Mihursky 1977; Holland, Shaughnessy, and Hiegel 1986) supported development of the long-term benthos monitoring program.

An important category of preliminary studies is the measurement of important parameters under different hydrological regimes to calibrate and validate water quality models. Such preliminary studies have been carried out in the New York Bight to assess the potential impacts of nutrient fluxes through the Sandy Hook transect (O'Connor, Mancini, and Guerriero 1981; Stoddard and Walsh 1988). Intensive monitoring in New York Harbor, especially during periods of high rainfall (Hazen and Sawyer and Hydroscience 1978), was used to develop mathematical models for a water quality management plan. Both projects provided the bases for development of larger-scale models for a water quality management plan of the combined New York-New Jersey Harbor and New York Bight. Similar studies, including both modeling and monitoring, have been performed in the James, Delaware, Patuxent, and Potomac estuaries. A significant feature of such studies is that monitoring focuses on variables used in modeling as well as on resources at risk.

These examples illustrate the fact that monitoring and research are interdependent, and that their frequent separation is more arbitrary than real. Rigidly separating monitoring and research decreases the chances that monitoring will benefit from new knowledge. Relevant research and other preliminary studies should be integrated throughout a monitoring program to resolve uncertainties and focus monitoring on test questions. Such an iterative sequence of planning, modeling, data collection, and research can result in highly effective monitoring programs. The experience in integrating these approaches is more extensive in freshwater environments (e.g., Cairns, Dickson, and Maki 1978).

STEP 4: DEVELOP SAMPLING/MEASUREMENT DESIGN

Linking Testable Questions to Useful Information

Step 4 uses the information produced in steps 1-3 to develop a sampling or measurement design that states what variables will be measured and where and when the measurements will be taken. (See Box 4.5.)

The steps outlined in Figure 4.6 ensure that sampling and measurement designs will be appropriate to the questions upon which the monitoring is based. The feedbacks ensure that the evolving sampling/measurement design will produce information needed to answer the specific questions to

BOX 4.5
DIFFERENT QUESTIONS, DIFFERENT SAMPLING DESIGNS

A simple example using sewage outfall effects will clarify the relationships between questions and the technical requirements of monitoring programs. If the focus is on differences in and outside the zone of initial dilution (ZID) boundary at one point in time, the sampling design will include stations in and outside the ZID at that one time. But if the focus is instead on changes over, say, five years, then these stations would be sampled periodically over the five years rather than just once. Year-to-year variability is not important in the first sampling design, but it is in the second.

be addressed. The elements of step 4 include: identifying the kinds and amounts of change that are meaningful; identifying and quantifying the sources of variability that may obscure or confound responses; deciding what to measure, in light of logistical constraints and limitations on scientific knowledge; developing a sampling design that provides the logical structure for the measurement program by specifying how variability will be partitioned; specifying statistical models that are the basis for selecting the kinds and numbers of measurements that should be taken; performing optimization and power analyses to determine whether the monitoring design can measure meaningful levels of change; defining data quality objectives; and developing the sampling/measurement design that incorporates all the above elements.

Defining Meaningful Change

The goal of a monitoring sampling design should be the detection of specific kinds and amounts of change in the resources at risk, in surrogate variables related to them, or in parameters involved in model validation or increasing the understanding of important natural processes (e.g., Fredette et al. 1986).

The definition of "meaningful change" is based on the testable questions developed in step 2. (See Box 4.6.) All kinds and levels of change are not equally important. It is therefore not possible to decide what parameters should be measured and when, where, and how measurements should be made until a determination is made about what kinds and levels of change are meaningful. When a decision about meaningful change has not been made, "monitoring programs run the risk, on the one hand, of

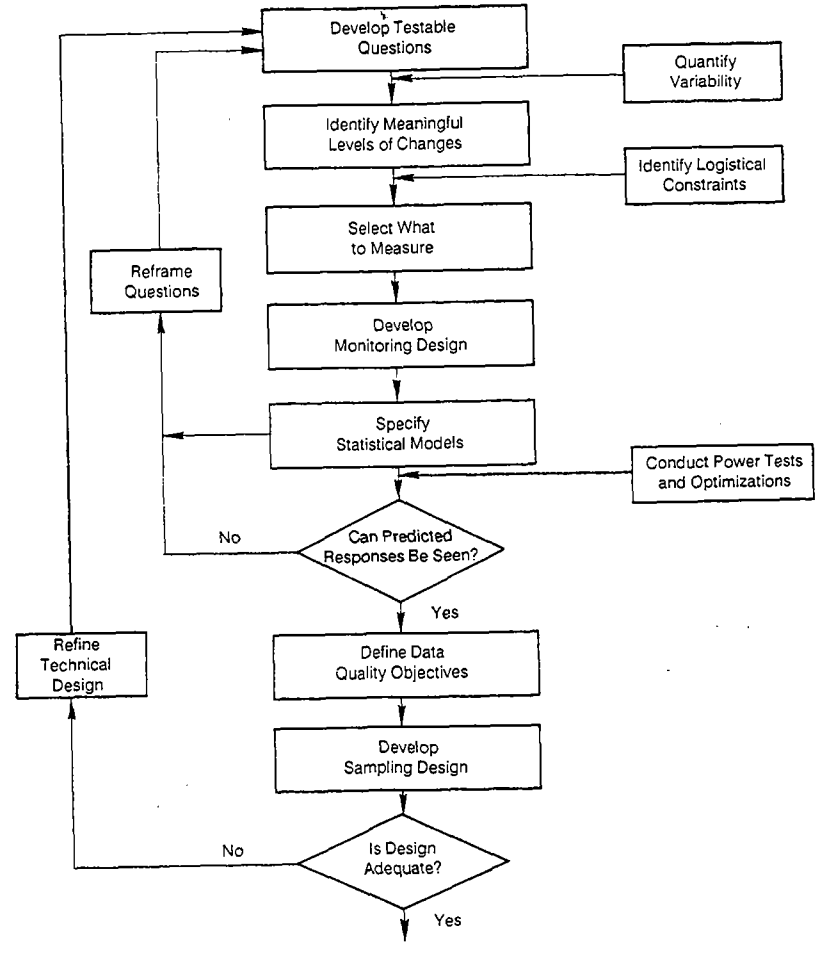

FIGURE 4.6 Step 4: Develop sampling/measurement design.

having little or no chance of detecting anything but catastrophic change, or, on the other, of sampling far in excess of what is necessary. . . ." (Bernstein and Zalinski 1983).

Deciding what kinds and amounts of change are meaningful (and to whom) is neither simple nor easy. Beanlands and Duinker (1983) note that statistical, scientific, project-specific, and wider societal concerns all contribute to the definition of meaningful changes. The benefits of defining how much change is meaningful cannot be overstated. This determination

BOX 4.6
MEANINGFUL VERSUS SIGNIFICANT CHANGE

Attempts to define meaningful amounts of change are confused by meanings of "significant" (Sharma, Buffington, and McFadden 1976; Christensen, Van Winkle, and Mattice 1976; Zar 1976). Significant means "having meaning . . .; having or likely to have influence or effect" (Merriam-Webster Inc. 1986), but it also refers to the statistical difference, at a specified probability level, between or among two or more sampling distributions. The California Ocean Plan (California State Water Resources Control Board 1983) prohibits activities causing "significant" change, defined in strictly statistical terms.

How large a change is important? One that is statistically significant is not necessarily meaningful. Virtually any change can be statistically significant, depending in part on the sampling effort. Thus a monitoring program with a small sampling effort will detect only large changes, but one with an intensive sampling effort could find even miniscule changes statistically significant. Whether changes in the environment are statistically significant has no bearing on the extent to which the changes may be either meaningful or important (i.e., have ecological or human consequences).

not only allows the designers of monitoring programs to focus resources more efficiently but also provides managers and decision makers with higher-quality information with which to make decisions. For example, the Marine Review Committee of the California Coastal Commission (1989) determined that an average decrease of 50 percent in population levels of certain plankton would indicate an impact from the San Onofore Nuclear Generating Station (SONGS) 3 kilometers away. A monitoring program was then designed with a specific probability of detecting this change. In the 106-mile dumpsite monitoring plan (Werme et al. 1988), the detection of any contaminants at the site boundary triggers another tier of monitoring activities. This decision criterion is based on the judgment that far-field effects are likely only if contaminants spread beyond the site boundary.

The definition of meaningful change is not static. It can shift with changing boundary conditions or new information. For example, a short-term one-time change in a water quality parameter or contaminant level should probably be viewed differently than the same degree of change in the long-term average.

The Influence of Natural Variability

Natural variability creates a background of change that may make it difficult to quantify environmental responses to human activity (Nichols 1985a). Thus defining meaningful change depends in part on identifying and accounting for natural sources of variability. For example, El Niños and occasional large winter storms in Southern California can destroy kelp bed canopies and prevent the detection of subtle impacts of human activities on the kelp beds. Seasonal changes in the abundance of the benthos in Chesapeake Bay affect population dynamics in ways that can also obscure benthic responses to human impacts (Holland, Shaughnessy, and Hiegel 1986). Similarly, large-scale climate-related shifts in fish distributions can make it difficult to identify and measure the effects of harvesting (Sherman and Alexander 1986).

Natural variation affects sampling design in two major ways. First, natural changes may be so large that they mask changes of human origin. Second, random or periodic variations not accounted for in the sampling design can result in noise or false signals that make it difficult to determine the response of the ecosystem (Christie 1985; Coull 1985; Lie and Evans 1973).

Understanding variability aids development of a sampling design in several ways: it helps construct a conceptual model that includes key natural processes and linkages that affect the resources being monitored, it helps partition variability by collecting data on appropriate space and time scales (Livingston 1987; Kerr and Neal 1976), and quantitative measures of variability provide input to the optimization and power analyses that predict whether the monitoring design can detect meaningful levels of change (Cohen 1988).

Characterizing variability can be difficult because of its many sources and scales in the marine environment. Natural spatial and temporal variability can reflect simple gradients in the physical environment (e.g., depth, salinity, and temperature), or it may reflect more complex processes such as succession and ecological interactions among ecosystem components (Levin 1978; Pearson and Rosenberg 1978; Nichols 1985b; Holland et al. 1986; Holling 1978). In addition to these natural kinds of variability, human activities and their impacts vary in space and time, and they can interact with natural processes to create intricate and sometimes perplexing patterns. (See Box 4.7.) As Wolfe et al. (1987) point out, cycles of temperature, light, and other factors interact with tidal cycles, seabed topography, and processes such as evaporation, turbulent diffusion, ion exchange, respiration, growth, and predation. Failure to understand fully such factors affecting the resources at risk can make it difficult or impossible to design monitoring programs that produce useful management information.

BOX 4.7
INTERACTIONS BETWEEN HUMAN ACTIVITY AND NATURAL PROCESSES

Impacts of human origin and natural processes sometimes vary together in ways that make determination of the causes of environmental change difficult. Events in the Southern California Bight between 1973 and 1977 are an example. There was a massive influx of the echiuran worm, *Listriolobus*, into benthic communities (Stull et al. 1986). This organism's burrowing, respiratory, and feeding activities aerated and reworked sediments throughout the bight. In areas of wastewater impacts (particularly around the White Point outfall in Los Angeles), these activities reduced apparent impacts. At the same time, mass emissions from the White Point outfall were substantially reduced. When the echiuran disappeared, impacts reappeared, but not to the extent seen previously. How much of the reduction in impacts was caused by the echiuran? How much by reduced discharges? It was hard, if not impossible, to say.

Each of the three uses of information about variability mentioned above (i.e., conceptual modeling, monitoring design development, and optimization and power analyses) requires somewhat different kinds of information. Building a realistic conceptual model requires a comprehensive review of all possible sources of variability. At this point in the monitoring design methodology, it is more important to have a qualitative understanding of the relationships among most or all sources of variability than a more quantitative description of a few. Failure to include an important source of variability can result in unrealistic assumptions about how impacts are created. For example, the lack of an atmospheric source term for nutrients in water quality models of Chesapeake Bay led to erroneous predictions and incomplete management strategies (Fisher et al. 1988; Tyler 1989).

Developing an adequate monitoring design depends on somewhat more quantitative knowledge about variability because the monitoring design must specify where and when measurements should be taken.

Optimization and power analyses require quantitative estimates of the major sources of variability. It is impossible to allocate limited sampling resources without such information. For example, if year-to-year variability in a particular system is much greater than seasonal variability, then proportionally more resources should be devoted to sampling additional years rather than additional seasons within years. Such decisions cannot be made

without knowledge about at least the relative magnitudes of the various sources of variability.

Selecting Variables to Measure

Most monitoring programs do not have the resources to monitor all variables of concern. The limited resources available must then be focused on the system attributes that are of the greatest concern and provide the most information about system status or changes in status. Thus actual sampling may not focus directly on the resources at risk but on surrogate variables. Surrogate variables include resources of intrinsic importance (e.g., economically important species, endangered species), early warning indicators (e.g., variables that respond rapidly to the stress of concern), sensitive indicators (e.g., variables that have a high degree of specificity to stress), process indicators (e.g., variables that provide insight into the effects of stress on complex system interactions, and variables with high information redundancy (i.e., those that are generally representative of the behavior of a number of important parameters). The rationale for monitoring surrogate variables is that they might provide clearer or simpler information than the resources would. This statement may not always apply (Wolfe and O'Connor 1986; O'Connor and Demling 1986; Bryan and Gibbs 1987), and specific criteria need to be applied to the selection of surrogate variables on a case-by-case basis. For example, diversity indices are often used to provide summary information about impacts on communities containing many species. However, much important information can be discarded in the calculation of these indices (May 1985). In addition, changes in diversity can be ambiguous, particularly when the study assemblage is exposed to more than one source of disturbance (NRC 1986). Criteria that should be used to select surrogate variables include sensitivity to the stress of concern, reliability and specificity of responses, ease and economy of measurements, and relevance of the indicator to specific concerns (NRC 1986).

Two important issues are involved in the choice of variables to monitor. The first relates to the depth of knowledge about a particular system (e.g., specificity and reliability of responses) and the second to the statistical efficiency of sampling alternative variables (e.g., the signal-to-noise ratio). A prime consideration for any monitored variable is that it should be tied directly to the specific questions to be answered and the resources at risk. In other words, changes in the status of the selected variable must unambiguously reflect changes in the resources at risk. How much they can be tied together depends largely on the depth of knowledge about the system and process being monitored. In well-understood systems, it will be clear which variables to measure and how to draw conclusions about the state of resources from them. For example, understanding the processes

BOX 4.8
VARIABILITY AFFECTS SELECTION OF VARIABLES

Dischargers in the Southern California Bight monitor the levels of contaminants in the tissue of fish collected around wastewater outfalls. But two potentially large and poorly understood sources of variability make it difficult to interpret these data. First, different species of fish are sampled at different outfalls (NRC in press). In other words, different variables (i.e., different species) are being sampled. Second, sampling is conducted at different times around the same outfall. However, contaminant levels in fish vary seasonally as a function of reproductive status (Cross et al. 1986). These two sources of variability may interact because of differences in the timing of reproductive cycles and in tissue chemistry among species, resulting in data that provide ambiguous information about the impacts of discharges on contaminant levels in fish or about the risk of contaminant discharge to the people who eat the fish.

leading to oxygen depletion and eutrophication has focused modeling and monitoring on nutrient levels (Hydroscience 1974; HydroQual 1986). When a system is less well understood, it may not be apparent which variables will indicate meaningful changes in resources. Then the conceptual model should be used to determine whether a particular variable can be linked to the specific questions to be answered with cause-effect statements. When crucial gaps in scientific understanding occur, research or modeling may be initiated to help determine what measurements should be made. In addition, the available information should be used to make an informed decision about what to monitor now. The kelp bed example described earlier (see Box 4.3) shows how research and modeling provided data that improved the conceptual model. This improved understanding was then used to focus monitoring on quantifying the response of kelp recruitment to power-plant-induced changes in near-bottom irradiance.

A second major consideration in selecting monitored variables is their statistical distributions and characteristics (e.g., signal-to-noise ratio). Monitored variables should provide the most accurate and precise estimates for the smallest required sampling effort, thus maximizing information return per sampling effort expended. Variables with high variability or unknown distributions (see Box 4.8) impair the ability to draw conclusions from monitoring data. Such variables are not appropriate for routine monitoring programs.

The Sampling Design and Its Statistical Basis

The sampling design is the central element in step 4 of the design methodology. (See Figure 4.6.) It provides the logical structure of the study (Cochran 1977; Fisher 1954) because it specifically defines how questions will be evaluated and how variation associated with different sources (e.g., spatial and temporal as well as human-induced variation) will be measured. For example, the kelp bed study (see Box 4.3) was structured around comparisons of characteristics of kelp beds located in the thermal plume against unaffected kelp beds located in reference areas far removed from the thermal plume. Several reference kelp beds were sampled to estimate natural variability among them. This structure defined the type of comparisons that would be used to detect impacts. In addition, the design consisted of sampling for several years before and several years after the power plant began operating to provide a background of natural temporal variability against which to measure changes in conditions that occurred once power plant operations began.

In many monitoring and assessment programs, it is not possible to collect preoperational data or to establish baseline conditions before an impact has occurred. Statistical comparisons in such cases are limited to comparing distributions among locations of concern to distributions at sites that are assumed to be appropriate reference areas (Green 1979). Selection of appropriate reference areas is always problematic. It is a particularly difficult problem in estuaries, where a natural salinity gradient that may vary in location from year to year generally requires broad regional sampling and application of estimation techniques to assess conditions that may occur at any particular location (Holland, Shaughnessy, and Hiegel 1986).

A poorly thought out sampling design usually results in testing of inappropriate questions, incomplete evaluation of questions, inability to separate change due to natural processes from change due to multiple activities, relatively low ability to detect change (low statistical power), and poor use of resources due to oversampling (e.g., Gore, Thomas, and Watson 1979; Hurlbert 1984; Stewart-Oaten and Murdoch 1986; Green 1979; Thomas 1978; Bernstein and Zalinski 1983; Toft and Shea 1983; Trautmann, McCulloch, and Oglesby 1982; Skalski and McKenzie 1982; Millard and Lettenmaier 1986). A well-planned sampling design, however, provides a logical basis for evaluating questions and a clear definition of a meaningful level of change, proper matching of variables with questions, quantification and partitioning of background variability, and proper assignment of sampling units among conditions or treatments.

Once a sampling design has been developed, it becomes the basis for a statistical model, which is a formal mathematical statement of the specific questions to be tested. By structuring how questions will be asked

and by formally describing and partitioning sources of variability, the statistical model furnishes an objective method for allocating sampling or measurement resources. Two statistical tools that aid in the fine-tuning and refinement of the sampling design are optimization and power analyses. When sampling resources are limited, optimization techniques help decide how to make trade-offs needed to control for several sources of variability (e.g., Gunnerson 1966). Power analysis is a procedure for determining the level of change a given sampling design will detect (Cohen 1988; Trautmann, McCullough, and Oglesby 1982). These analyses can be conducted before samples are taken, after part of the samples have been collected, or after the program has ended. This knowledge can be invaluable in determining whether the resources available for monitoring are likely to produce useful information before a program is initiated. If power analyses show that meaningful levels of change cannot be detected with the available resources, then the monitoring program can be redirected before these resources are wasted on trying to answer unanswerable questions. They also provide scientists and decision makers with an estimate of the level of uncertainty and thus the degree of confidence they should place in a given analysis result at the conclusion of a program.

QUALITY ASSURANCE: AN IMPORTANT ELEMENT OF MONITORING PROGRAM DESIGN AND IMPLEMENTATION

A quality assurance program is a system of activities undertaken to ensure that the type, amount, and quality of data collected are adequate to meet study objectives; it is a critical element of all monitoring programs (Taylor 1985; EPA 1979; EPA 1984a). Quality assurance consists of two separate but interrelated activities: quality control and quality assessment (Taylor 1985).

Quality control includes activities to ensure that the data collected are of adequate quality given study objectives and the specific hypotheses to be tested (steps 1-4). Quality control activities frequently undertaken within monitoring programs include standardized sample collection and processing protocols and requirements for technician training (EPA 1984b). The goals of quality control procedures are to ensure that sampling, processing, and analysis techniques are applied consistently and correctly; the number of lost, damaged, and uncollected samples is minimized; the integrity of the data record is maintained and documented from sample collection to entry into the data record; the data are comparable with similar data collected elsewhere; and study results can be reproduced.

Quality assessment activities are implemented to quantify the effectiveness of the quality control procedures. They ensure that measurement error is estimated and accounted for and that bias associated with the monitoring

program can be identified and, if practical, eliminated. Quality assessment consists of both internal and external checks, including repetitive measurements, internal test samples, interchange of technicians and equipment, use of independent methods to verify findings, exchange of samples among laboratories, use of standard reference materials, and audits (Taylor 1985; EPA 1980, 1984c).

To be effective, quality assurance must begin with planning the monitoring program. Thus the level of uncertainty associated with obtaining the required information can be balanced against the cost of obtaining the data (EPA 1984b). Steps 1-5 activities for defining what to measure and how, where, and when to take measurements are all part of the quality assurance process. Quality assurance must continue to be an integral component of monitoring systems from implementation through information dissemination. Activities for converting the data into useful information (steps 6-7) and the feedback loops shown in Figure 4.1 must also be taken into account in designing the quality assurance program. These later activities provide mechanisms for using quality assessment information to modify and improve monitoring.

The need for quality assurance programs increases with the complexity of the measurement program and the number of organizations involved (Taylor 1978, 1985). Experience shows that chemical monitoring programs that involve a number of laboratories measuring concentrations of chemical substances are particularly subject to quality assurance problems (Taylor 1985). For example, during the early stages of the Chesapeake Bay Monitoring Program, nutrient data were collected and analyzed by three regional laboratories, all using different protocols for processing samples. As a result, the data were not comparable and could not be used to depict nutrient distributions accurately (Martin Marietta Environmental Systems 1987). As is often the case, because of the haste to initiate the collection program, the laboratories' methods and equipment were not evaluated (Taylor 1985).

Another important quality assurance issue associated with monitoring systems is maintaining the integrity of large data sets (Packard, Guggenheim, and Bernstein 1989). Two general data management problems must usually be resolved: (1) correction or removal of erroneous individual values and (2) inconsistencies that damage the integrity of the data base. Many erroneous individual values can be identified, validated, and corrected using range checks, filtering algorithms, and comparison to lists of valid values. Entering data twice using different data entry operations and then checking for nonmatches are a particularly effective method for identifying and correcting key-punch errors. Subtle errors that affect the integrity of multiple data entries are much more difficult to identify and correct. For example, errors that affect the relationships among data entries are particularly difficult to identify and correct, especially in large regional monitoring

data bases. Although some data base management systems protect against such errors, others require rigorous cross-checking during data entry to identify and correct these errors. Experience shows that the most effective way to avoid corruption of a data base is to select a data management system that protects against internal inconsistencies and to design the data entry process to minimize the occurrence of errors (Packard, Guggenheim, and Bernstein 1989). Data entry screens should be simple, and they should mimic the layout of raw data sheets. Typographical errors can be minimized by users selecting from a list of valid values (using lookup tables) rather than typing in the actual values.

Quality assurance activities ensure that the goals and objectives of the monitoring program are achieved and that the data that result are adequate for use in making the anticipated decisions. The final and perhaps most important component of quality assurance for a monitoring system is the external review process. Expert reviews should be conducted before samples are taken, at various logical interim phases during a program, and following the analysis and interpretation of the data.

STEP 6: CONVERT DATA INTO USEFUL INFORMATION

The raw data collected in a monitoring program frequently do not directly address public concerns or the information needs of decision makers. Data are individual facts, and information is data that have been processed, synthesized, and organized for a specific purpose. Drucker (1988) described the difference between data and information: "Information is data endowed with relevance and purpose. Converting data into information thus requires knowledge." A useful monitoring program provides knowledge or, more specifically, mechanisms to ensure that knowledge is used to convert data collected into information.

For example, measurements of contaminant concentrations in the water or sediments near a discharge in and of themselves are not useful information. Contaminant concentration data must be analyzed and mapped to describe patterns and trends. They must then be combined with additional data (e.g., background levels, transport processes, and flux rates) to define exposure. Ultimately, to assess environmental impacts, exposure information must then be combined with the results of studies of pollutant transport and effects research (e.g., bioassay experiments) to assess the risks to and consequences for receptors and processes. Conversion of monitoring data into information, therefore, involves a range of activities, including data management, statistical analysis, predictive modeling, and fate and effects research. Each of these activities is discussed below.

Data Management

The major function of data management activites is to provide easy access to the collected data and related information (e.g., historical trends data, research data, model outputs, data summaries). Because of the amount and complexity of the data that are collected by most monitoring programs and the variety of reports and analyses that are produced, a computer-assisted data management system is usually essential. To define and select the appropriate data management system, managers should first determine the volume of data, the long-term uses of the data, existing data management capabilities, the number and background of and relationships among users of the data, the major types of analyses to be conducted, and quality assurance/quality control and reporting requirements. This information ensures a system with the required capacity and degree of access.

Monitoring data can be stored in a central location. They can also be accessed through a distributed data management system. In either case, monitoring data and relevant model results should be included in both raw and summarized form to eliminate costly reanalysis. In addition, information on study characteristics, information on the institution responsible for data collection and storage, and a brief description of sampling methods, data format, quality control procedures, and how to access the data should be readily available for each data set.

Data management activities are as important to the success of monitoring programs as the collection of data. Therefore they should be funded as a continuing core program element, and reports that summarize the types, volume, and quality of data accessible through the system should be prepared and distributed to potential users frequently. Unfortunately, monitoring data are frequently not incorporated into a data management system until most data collection is complete. At this point in many programs, there may not be enough time or money to create an adequate system. This situation lessens the utility of monitoring data to scientists within and outside the program.

Data Analysis and Modeling

The goals of analysis activities are to summarize and simplify the collected data, test for change and differences, generate hypotheses, determine the consequences of observations, and evaluate the uncertainty associated with conclusions drawn from the data. Analysis programs should be developed prior to data collection. This development should include both statistical testing and modeling to ensure that the analysis approach is appropriate to the sampling design and the sampling methods.

Successful analysis programs cut across institutional and media boundaries; partition spatial and temporal variations into their major sources (natural and human induced); are based on an understanding of linkages among physical, chemical, and biological attributes; use standard verified modeling approaches, statistical packages, and analysis/data management packages; state and determine the consequences of assumptions inherent in the sampling design and analysis approach; evaluate the sensitivity of analyses to assumptions; and summarize analysis results using easily understood graphs, maps, and tables.

Statistical analysis helps characterize the data, determine the uncertainty associated with measurements, classify the data into appropriate spatial and temporal strata, and test for spatial and temporal change. Generally, many statistical tests are appropriate for any particular situation. Selection of the most appropriate test depends upon data characteristics and the specific question being asked. Numerous publications are available to help scientists and nonscientists identify the most appropriate test, conduct the test, and interpret the results (e.g., Green 1979).

As discussed earlier, forecasting the responses of complex marine systems to human activities and assessing their status and trends with reliability are a difficult problem. Simulation models are an assessment tool that can be used to describe environmental complexities while allowing these complexities to be used in forecasting the consequences of environmental change. Simulation models are based on essential system attributes.

Research is a basic element in the development of predictive models and the interpretation and synthesis of monitoring data and model outputs. It is the major process for establishing cause-effect relationships. Correlations and relationships identified during the analysis of monitoring data (e.g., Cairns, Dickson, and Maki 1978; Smith, Bernstein, and Cimberg 1988; Holland, Shaughnessy, and Hiegel 1986) can be an important source of ideas for future experiments and measurements. The Southern California Bight case study found that monitoring programs had benefited greatly from their close association with ongoing research programs designed to understand the fate of discharged wastes and assess sublethal effects. The Southern California experience also shows that the results from separately managed and funded research programs can be transferred effectively.

Resource allocations for analysis activities are frequently not commensurate with those for data collection. For example, the Chesapeake Bay case study found that far too little attention and resources were directed at data analysis and synthesis relative to the investment made to collect the data. Data should not be collected unless a commitment is made at the outset that support for analysis activities will be commensurate with that for data collection.

One way to address the above problem is to use a phased analysis approach. In such an approach, the data collected early in the monitoring program are used to develop and refine routine analysis methods, classify the data into spatial and temporal components, determine the adequacy of the sampling design and methods, define the status and its relationship to historical conditions, and develop a preliminary understanding of links between components and processes. Interdisciplinary analyses can follow later in the program.

STEP 7: DISSEMINATE RESULTS

The results of monitoring programs, especially regional programs, should be disseminated to a range of audiences and at several technical levels. Monitoring programs that produce only technical reports summarizing data and scientific findings are not likely to show the public or decision makers that they provide information essential to better environmental protection or management decisions. In fact, management information is produced only when it is delivered to managers and decision makers in a usable, accessible form. Many monitoring programs, especially status and trends studies, extend over years. Interim results of these studies should be disseminated regularly, allowing users to determine whether the type and volume of data that they need are being obtained. If the needed information is not being obtained, midcourse adjustments can then be made. A phased analysis and reporting approach similar to that used by the Maryland Department of the Environment (see Box 4.9) keeps target audiences informed about what the information being collected means, what data remain to be collected, what analyses remain to be completed, and why additional data collection and analyses are needed.

REALISTIC EXPECTATIONS

While acknowledging the importance and utility of monitoring information, one must not overstate the utility of monitoring information. The marine environment is complex and variable, and it is often difficult to identify and measure clearly the impacts of human origin. These factors, coupled with limitations to scientific knowledge, emphasize the need for realistic expectations. Management of the environment and the monitoring programs that are a part of that management must therefore consider the risks and uncertainties inherent in most actions. Monitoring is limited in terms of its ability to quantify changes and to identify their causes. These limitations must be forthrightly stated, understood, and incorporated in the decision-making process.

The reality of imperfect knowledge about marine systems means that monitoring should be used as an opportunity to increase and refine our

BOX 4.9
DISSEMINATION OF INFORMATION IN THE CHESAPEAKE BAY PROGRAM

The Maryland Department of the Environment (MDE) Chesapeake Bay Water Quality Monitoring Program was designed to assess water quality conditions for the Maryland Chesapeake Bay and to determine the effectiveness of actions and policies to improve and protect water quality. The program disseminates its results to the public, scientists, and decision makers. The reports described here are an example of what monitoring programs should produce.

Level I Reports

Level I reports, prepared semiannually, summarize the status of data collection activities; they include displays of spatial, seasonal, and long-term trends, analyses of results, and tabular data summaries. One of the two reports also summarizes analyses. They are distributed to all appropriate agencies and organizations.

Level II Reports

Level II reports, prepared every two years, reach the same audience as Level I reports, but they are more interpretive. Level II reports evaluate relationships among study elements, place the data in an ecological and regional perspective, and quantify the effects of major processes affecting water quality.

Level III Reports

These reports are prepared periodically for politicians, high-level decision makers, and the public. They provide an overall assessment of the status of Chesapeake Bay and changes that have occurred over defined periods. Their objectives are to identify the factors influencing environmental conditions, evaluate restoration actions, and identify management actions and policies that would improve conditions.

Executive Summaries

Program summaries, prepared annually, are short documents prepared for each major program element. They list the data being collected; describe how, when, and where collections are made; list the name, telephone number, organization, and address of the responsible principal investigator(s); describe how to obtain data summaries and/or raw data; highlight major findings, conclusions, and recommendations; and describe future plans.

Additional Documents

Periodically, MDE prepares and disseminates field and laboratory manuals, data management reports, and findings of special studies conducted to evaluate sampling and processing methods.

knowledge of them. Data and information derived from monitoring programs should be used to check, validate, and refine the assumptions, models, and understandings on which the monitoring was based. This iterative feedback increases predictive ability, reduces uncertainty, and ultimately reduces the monitoring effort needed. As discussed in Chapter 2, risk-free decision making is not achievable, and monitoring must be viewed as a way of reducing uncertainty, not of eliminating it.

Although not a necessary ingredient of every monitoring program, research on natural variability and its causes, ecosystem function, transport and fate of materials, and biological effects of contaminants and habitat alterations is critical to the evolution of knowledge that makes monitoring more effective. At the least, regional trends monitoring should be accompanied by an ongoing research program designed to contribute to the interpretation of monitoring results. If it is not, the accumulation of data will outstrip maximum use of these data or, worse, will lead to erroneous conclusions.

In most monitoring efforts, the need to hold study methods constant for the sake of continuity must be balanced against the need to adapt methods to reflect technological advances. This dilemma cannot be resolved in any arbitrary fashion, and it must be carefully and periodically addressed in each monitoring program. Such adaptation not only includes the collection of additional data and application of new sampling techniques, but it also includes dropping obsolete measurements, reducing monitoring efforts for well-understood processes, and restructuring the entire program when fundamental assumptions are found to be flawed. As knowledge improves and new problems come to light, the resources available for monitoring must be shifted appropriately. Thus a crucial part of technical design is knowing when to stop or reduce the monitoring effort devoted to a particular problem.