United States Environmental Protection Agency
Office of Water
Washington, DC 20460
EPA 822-R-02-015
March 2002

METHODS FOR EVALUATING WETLAND CONDITION

#4 Study Design for Monitoring Wetlands

United States Environmental Protection Agency
Office of Water
Washington, DC 20460
EPA 822-R-02-015
March 2002

METHODS FOR EVALUATING WETLAND CONDITION

#4 Study Design for Monitoring Wetlands

Principal Contributor

Warnell School of Forest Resources, University of Georgia
Amanda K. Parker

Prepared jointly by:

The U.S. Environmental Protection Agency
Health and Ecological Criteria Division (Office of Science and Technology)

and

Wetlands Division (Office of Wetlands, Oceans, and Watersheds)

Notice

The material in this document has been subjected to U.S. Environmental Protection Agency (EPA) technical review and has been approved for publication as an EPA document. The information contained herein is offered to the reader as a review of the “state of the science” concerning wetland bioassessment and nutrient enrichment and is not intended to be prescriptive guidance or firm advice. Mention of trade names, products or services does not convey, and should not be interpreted as conveying official EPA approval, endorsement, or recommendation.

Appropriate Citation

U.S. EPA. 2002. *Methods for Evaluating Wetland Condition: Study Design for Monitoring Wetlands.* Office of Water, U.S. Environmental Protection Agency, Washington, DC. EPA-822-R-02-015.

This entire document can be downloaded from the following U.S. EPA websites:

http://www.epa.gov/ost/standards

http://www.epa.gov/owow/wetlands/bawwg

Contents

List of Tables

Foreword

In 1999, the U.S. Environmental Protection Agency (EPA) began work on this series of reports entitled *Methods for Evaluating Wetland Condition*. The purpose of these reports is to help States and Tribes develop methods to evaluate (1) the overall ecological condition of wetlands using biological assessments and (2) nutrient enrichment of wetlands, which is one of the primary stressors damaging wetlands in many parts of the country. This information is intended to serve as a starting point for States and Tribes to eventually establish biological and nutrient water quality criteria specifically refined for wetlands.

This purpose was to be accomplished by providing a series of "state of the science" modules concerning wetland bioassessment as well as the nutrient enrichment of wetlands. The individual module format was used instead of one large publication to facilitate the addition of other reports as wetland science progresses and wetlands are further incorporated into water quality programs. Also, this modular approach allows EPA to revise reports without having to reprint them all. A list of the inaugural set of 20 modules can be found at the end of this section.

This series of reports is the product of a collaborative effort between EPA's Health and Ecological Criteria Division of the Office of Science and Technology (OST) and the Wetlands Division of the Office of Wetlands, Oceans and Watersheds (OWOW). The reports were initiated with the support and oversight of Thomas J. Danielson (OWOW), Amanda K. Parker and Susan K. Jackson (OST), and seen to completion by Douglas G. Hoskins (OWOW) and Ifeyinwa F. Davis (OST). EPA relied heavily on the input, recommendations, and energy of several panels of experts, which unfortunately have too many members to list individually:

- Biological Assessment of Wetlands Workgroup
- Wetlands Nutrient Criteria Workgroup

More information about biological and nutrient criteria is available at the following EPA website:

http://www.epa.gov/ost/standards

More information about wetland biological assessments is available at the following EPA website:

http://www.epa.gov/owow/wetlands/bawwg

List of "Methods for Evaluating Wetland Condition" Modules

Summary

State and Tribal monitoring programs should be designed to assess wetland condition with statistical rigor while maximizing available management resources. The three study designs described in this module—stratified random sampling, targeted/tiered approach, and before/after, control/impact (BACI)—allow for collection of a significant amount of information for statistical analyses with relatively minimal effort. The sampling design selected for a monitoring program will depend on the management question being asked. Sampling efforts should be designed to collect information that will answer management questions in a way that will allow robust statistical analysis. In addition, site selection, characterization of reference sites or systems, and identification of appropriate index periods are all of particular concern when selecting an appropriate sampling design. Careful selection of sampling design will allow the best use of financial resources and will result in the collection of high-quality data for evaluation of the wetland resources of a State or Tribe. Examples of different sampling designs currently in use for State and Tribal wetland monitoring are described in the Case Study (Bioassessment) module and on http://www.epa.gov/owow/wetlands/bawwg/case.html.

Purpose

The purpose of this module is to provide technical guidance information on designing effective sampling programs for State and Tribal wetland water quality monitoring.

Introduction

Wetlands are included as waters of the United States in the Federal regulations (40 CFR 122.2, 40 CFR 230.3, and 40 CFR 232.2) implementing the Clean Water Act [Section 502(7)]. Wetlands are important waterbodies; they can provide many functions that are beneficial to the local landscape, for example, water storage, water quality improvement, and wildlife habitat. Wetlands are also valuable as ecosystems in their own right, providing carbon storage, biogeochemical transformations, and aquifer recharge (Mitsch and Gosselink 1993). However, few States or Tribes (only six States and Tribes reported attainment of designated uses in the "National Water Quality Inventory 1996 Report to Congress") include wetland monitoring in their routine water quality monitoring programs (U.S. EPA 1996, 1998). Yet, wetland chemical and biological water quality monitoring data are scarce to nonexistent in many States and most Federal databases. The need for water quality monitoring data on wetlands is obvious; the vast majority of data about wetlands are collected in relation to dredge and fill permitting. Indeed, only 4% of the nation's wetlands were surveyed and only 11 States and Tribes reported information concerning wetland designated use attainment in the "National Water Quality Inventory 1998 Report to Congress" (U.S. EPA 1998). This module is intended to provide guidance to State and Tribal water quality managers on designing wetland monitoring programs to be included as a part of their routine water quality sampling.

Most States and Tribes will need to begin wetland monitoring programs to collect water quality and biological data (U.S. EPA 1990). The best monitoring programs are designed to assess wetland condition with statistical rigor while maximizing available management resources. At the broadest level, monitoring should include the following:

- Detecting and characterizing the ambient condition of existing wetlands
- Describing whether wetland condition is improving, degrading, or staying the same
- Defining seasonal patterns in wetland conditions
- Identifying thresholds for system stressors, that is, how much the system can be disturbed with-

out causing unacceptable changes in wetland system quality or degradation of beneficial uses.

Water quality monitoring programs at the State and Tribal level are often poorly and inconsistently funded or are improperly designed and carried out, making it difficult to collect a sufficient number of samples over time and space to identify changes in system condition or to estimate average conditions with statistical rigor. Three approaches to study design for assessing water quality as well as biological and ecological condition and for identifying degradation in wetlands are described in this module. Specific issues to consider in designing monitoring programs for wetland systems are also discussed in this module. The study designs presented can be tailored to fit the specific goals of monitoring programs. The assemblage-specific modules will discuss detailed sampling considerations.

The three approaches described—stratified random sampling; targeted/tiered approach; and BACI—present study designs that allow one to obtain a significant amount of information with relatively minimal effort. Stratified random sampling begins with a large-scale random monitoring design that is reduced as the wetland system or habitat conditions are characterized. This approach is used to find the mean condition of each wetland class, or type, in a specific region. Stratified random sampling design is frequently used for new large-scale monitoring programs at the State and Federal level (e.g., the Environmental Monitoring and Assessment Program, the Regional Environmental Monitoring and Assessment Program, and State programs in Maine, Montana, and Wisconsin). The tiered or targeted approach to monitoring begins with coarse screening and proceeds to more detailed monitoring protocols as impaired and high-risk systems are identified for further investigation. Targeted sampling design provides a triage approach to identifying wetland systems in need of restoration, protection, and intensive management. Several State pilot projects use this method or a modification of this method for wetland assessment (e.g., Florida, Ohio, Oregon, and Minnesota). The synoptic approach described in Kentula et al. (1993) uses a modified targeted sampling design. The BACI design and its modifications are frequently used to assess the success of restoration efforts or other management experiments. BACI design allows for comparisons in similar systems over time to determine the rate of change in relation to the management activity, for example, to assess the success of a wetland hydrologic restoration. Detenbeck et al. (1996) used BACI design for monitoring water quality of wetlands in the Minneapolis/St. Paul, Minnesota, metro area.

Monitoring programs should be designed to answer questions such as how, when, where, and at what levels do unacceptable wetland conditions occur? These questions are interrelated, and a well-designed monitoring program can contribute to answering them. Sampling design is dependent on the management question being asked. Sampling efforts should be designed to collect information that will answer the management question. For example, stratified random sampling might be good for ambient monitoring programs, BACI for evaluating restoration, and targeted sampling for developing an Index of Biological Integrity or nutrient criteria thresholds. In fact, some State programs will likely need to use a combination of approaches.

Considerations for Sampling Design

Describing the Management Question

Clearly defining the question being asked (identifying the hypothesis) encourages the use of appropriate statistical analyses, reduces the occurrence of false-positive (Type I) errors, and increases the efficient use of management resources (Suter 1993, Leibowitz et al.1992, Kentula et al. 1993). Beginning a study or monitoring program with carefully defined questions and objectives helps to identify the statistical analyses most appropriate for the study and reduces the chance that statistical assumptions will be violated. Management resources are optimized because resources are directed at monitoring that is most likely to answer management questions. In addition, defining the specific hypotheses to be tested, carefully selecting reference sites, and identifying the most useful sampling interval can help reduce the uncertainty associated with the results of any sampling design as well as further conserve management resources (Kentula et al. 1993).

Site Selection

Site selection is arguably the most important task in developing a monitoring program (Kentula et al. 1993). Site selection for a monitoring program is based on the need to sample a relatively large number of wetlands to establish the range of wetland quality in a specific regional setting. Protecting or improving the quality of a wetland system often depends on the ability of the monitoring program to identify dose-response relationships, for example, the relationship of nutrient concentration (dose) to periphyton abundance (response). Dose-response relationships can be identified using large sample sizes and systems that span the gradient (low to high) of wetland quality. All ranges of responses should be observed along the dosing gradient from low levels to high levels of human disturbance. In addition, wetland monitoring frequently requires an analysis of both watershed/landscape characteristics and wetland-specific characteristics (Kentula et al. 1993, Leibowitz et al. 1992). Therefore, wetland sampling sites should be selected on the basis of land use in the region so that watersheds range from minimally impaired with few expected stressors to high levels of development (e.g., agriculture, forestry, or urban) with multiple expected stressors (see Module 17: Land-Use Characterization for Nutrient and Sediment Risk Assessment). Establishing dose-response relationships may be confounded due to time lags between stressor occurrence and biological or functional response. The duration of the time lag between stressor and response depends on many factors, including the type of stressor, climate, and system hydrology. These factors should be considered when selecting wetland sites to establish the range of wetland quality within a region.

The synoptic approach described in Leibowitz et al. (1992) provides a method of rapid assessment of wetlands at the regional and watershed levels that can help identify the range of wetland quality within a region. Leibowitz et al. (1992) recommend an initial assessment for site selection based on current knowledge of watershed- and landscape-level features; modification of such an assessment can be made as more data are collected. Assessing watershed characteristics through the use of aerial photography and geographical information systems linked to natural resource and land-use databases can aid in identifying reference and degraded systems (see Module 17: Land-Use Characterization for Nutrient and Sediment Risk Assessment) (Johnston et al. 1988, 1990, Gwin et al. 1999, Palik et al. 2000). Some examples of watershed characteristics that can be evaluated using aerial photography and geographical information systems include land use, land cover (including riparian vegetation), soil, bedrock, hydrography, and infrastructure (e.g., roads and railroads).

Identifying and Characterizing Reference wetlands

The term "reference" in this module refers to those systems that are least impaired by anthropogenic effects. This term can be confusing because of the different meanings that are currently in use in different classification methods (particularly in hydrogeomorphic wetland classification). A discussion of the term "reference" and its multiple meanings is provided in Module 7: Wetlands Classification.

Watersheds with little or no development that receive minimal anthropogenic inputs could potentially contain wetlands that would serve as minimally impaired reference sites. Watersheds with a high percentage of the drainage basin occupied by urban areas, agricultural land, and altered hydrology are likely to contain wetlands that are impaired or could potentially be considered "at risk" for developing problems. Wetland loss in the landscape should also be considered when assessing watershed characteristics for reference wetland identification. Biodiversity can become impoverished due to wetland fragmentation or decreases in regional wetland density, even in the absence of site-specific land-use activities. Reference wetlands may be more difficult to locate if fragmentation of wetland habitats is significant, and they may no longer represent the biodiversity of minimally disturbed wetlands in the region. The continued high rate of wetland loss in most States and Tribes requires that multiple reference sites be selected to ensure some consistency in reference sites for multiple-year sampling programs (Leibowitz et al. 1992, Kentula et al. 1993). Once the watershed level has been considered, a more site-specific investigation can be initiated to better assess wetland condition.

Potential reference wetlands should be characterized to allow for the identification of appropriate reference wetland systems. Appropriate reference sites will have similar soils, vegetation, hydrologic regimes, and landscape settings to other wetlands in the region (Adamus 1992, Leibowitz et al. 1992, Kentula et al. 1992, Detenbeck et al. 1996). Classification of wetlands, as discussed in Module 7: Wetlands Classification, will aid in identifying appropriate reference wetlands for specific regions and wetland classes. Wetland classification should be supplemented with information on hydroperiod and flood frequency to ensure that the selected reference wetlands are truly representative of wetlands in the region, class, or subclass of interest. Reference wetlands may not be available for all wetland classes. In this case, data from systems that are as close as possible to the assumed unimpaired state of wetlands in the wetland class of interest should be sought from within the same geologic province. Development of a conceptual reference may be necessary if appropriate reference sites cannot be found in the local region or geologic province. Techniques for defining a conceptual reference are discussed at length in Harris et al. (1995), Trexler (1995), and Toth et al. (1995).

Reference wetlands should be selected on the basis of low levels of human alteration in their watersheds (Leibowitz et al. 1992, Kentula et al. 1993, U.S. EPA 2000). Selecting reference wetlands usually involves the assessment of land use within watersheds as well as visits to individual wetland systems to ground-truth expected land use and check for unsuspected impacts. Ground-truthing visits to reference wetlands are crucial for the identification of ecological impairment that may not be apparent from land-use and local habitat conditions. Again, a sufficient sample size is important to characterize the range of conditions that can be expected in the least impacted systems of the region (Detenbeck et al. 1996). Reference wetlands should be identified for each ecoregion or geological province in the State or Tribal lands and then characterized with respect to ecological integrity. A minimum of three low-impact reference systems is recommended for each wetland class for statistical analyses. However, power analysis can be performed to determine the degree of replication necessary to detect an impact to the systems being investigated (Detenbeck et al. 1996, Urquhart et al. 1998; see also http://www.mp1-pwrc.usgs.gov/powcase/index.html). Highest priority should be given to identifying reference systems for those wetland classes considered to be at the greatest risk from anthropogenic stress.

When to sample

Sampling may be targeted to the period when problems are most likely to occur—the index period. The appropriate index period will be defined by what the investigator is trying to investigate and by what taxonomic assemblage or parameters are being used for that investigation (Barbour et al. 1999). For example, increased nutrient concentrations and sedimentation from nonpoint sources may occur following periods of high runoff during the spring and fall, whereas point sources of nutrient pollutants may cause plankton blooms and/or increased water and soil nutrient concentrations in wetland pools during times of low rainfall. Hence, different index periods may be needed for nonpoint source and point source nutrients. Each taxonomic assemblage studied will also have an appropriate index period—usually in the growing season (see assemblage methods in the Minnesota case study: http://www.epa.gov/owow/wetlands/bawwg/case/mn1.html). The index period window may be early in the growing season for amphibians and algae. Other assemblages, such as vegetation and birds, may require a different sampling window for the index period (see the assemblage-specific modules for recommendations). Once wetland condition has been characterized, one-time annual sampling during the appropriate index period may be adequate for multiple-year monitoring of indicators of nutrient status, designated use, and biotic integrity. However, criteria and ecological indicator development may require more frequent sampling to define conditions that relate to the stressor or the impact of interest (Karr and Chu 1999, Stevenson 1996, 1997).

Ideally, water quality monitoring programs produce long-term data sets compiled over multiple years to capture the natural, seasonal, and year-to-year variations in biological communities and waterbody constituent concentrations (Tate 1990, Dodds et al. 1997, McCormick et al. 1999). Multiple-year data sets can be analyzed with statistical rigor to identify the effects of seasonality and variable hydrology. Once the pattern of natural variation has been described, the data can be analyzed to determine the ecological state of the waterbody. Long-term data sets have also been important in influencing management decisions about wetlands, most notably in the Everglades, where long-term data sets have induced Federal, State, and Tribal actions for conservation and restoration of the largest wetland system in the United States (see Davis and Ogden 1994, Redfield 1999, 2000, 2001; 1994 Everglades Forever Act, Florida Statute § 373.4592).

In spite of the documented value of long-term data sets, there is a tendency to intensively study a waterbody for 1 year before and 1 year after treatment. A more cost-effective approach would be to measure only the indices most directly related to the stressor of interest (i.e., those parameters or indicators that provide the best information to answer the specific management question), but to double or triple the monitoring period. Two or more years of data are often needed to identify the effects of years with extreme climatic or hydrologic conditions. Comparisons over time between reference and at-risk or degraded systems can help describe biological response and annual patterns in the presence of changing climatic conditions. Long-term data sets can also help describe regional trends. Flooding or drought may significantly affect wetland biological communities and the concentration of water column and soil constituents. Effects of uncommon climatic events can be characterized to discern the overall effect of management actions (e.g., nutrient reduction and water diversion) if several years of data are available to identify the long-term trends. At the very minimum, 2 years of data before and 2 after specific management actions, but preferably 3 or more each, are recommended to evaluate the cost-effectiveness of management actions with some degree of certainty (U.S. EPA 2000). If funds are limited, restricting sampling frequency and/or numbers of indices analyzed should be considered to preserve a long-term data set. Reducing sampling frequency or numbers of parameters measured will allow for effectiveness of management approaches to be assessed against the high annual variability that is common in most wetland systems. Wetlands with high hydrological variation from year to year may require more years of sampling before and after mitigation procedures to identify the effects of the natural hydrologic variability (Kadlec and Knight 1996).

Using the BACI study design may also provide substantial benefit for determining the effectiveness of management activities. Tracking both reference (control) and impacted systems within a region over time will help determine the rate and direction of change in monitored systems regardless of system variability. The BACI design may require more frequent monitoring of reference systems than is suggested for random or targeted designs, but it would allow useful comparison of system change regardless of variability (Detenbeck et al. 1996).

Characterizing precision of estimates

Estimates of dose-response relationships, nutrient and biological conditions in reference systems, and wetland conditions in a region are based on sampling, hence precision must be assessed. Precision is defined as "measure of the degree of agreement among the replicate analyses of a sample, usually expressed as the standard deviation" (Eaton et al. 2000). Determining precision of measurements for one-time assessments from single samples in a wetland is often necessary. The variation associated with one-time assessments from single samples can often be determined by resampling a specific number of wetlands during the survey. Measurement variation among replicate samples can then be used to establish the expected variation for one-time assessment of single samples. Resampling does not establish the precision of the assessment process; it identifies the precision of an individual measurement (Kentula et al. 1993). Resampling frequency is often conducted for one wetland site in every block of 10 sites. However, investigators should adhere to the objectives of resampling (often considered an essential element of quality assessment/quality control) to evaluate the variation in a one-time assessment from a single sample. The larger the sample size, the better (smaller) will be the estimate of that variation. Often, more than 1 in 10 samples need to be replicated in monitoring programs to provide a reliable estimate of measurement precision (Barbour et al. 1999).

Sampling Protocol

Approaches to Sampling Design

Three approaches to sampling design—stratified random, targeted design, and BACI—have advantages and disadvantages that under different circumstances warrant the choice of one approach over the other (Table 1). The decision as to the best approach for sample design in a new monitoring program must be made by the water quality resource manager or management team after carefully considering different approaches. Justification of a dose-response relationship is confounded by lack of randomization and replication, and must be considered in choosing a sampling design for a monitoring program. Direct identification of a cause-response relationship is not possible in observational (monitoring) studies. However, inferences of causality can be argued if appropriate information is collected. Beyers (1998) describes assembly rules for causal arguments that can be used to infer causality for stressors of interest. The number of sites to be sampled and the sampling frequency are determined by the type of sampling design chosen by the resource managers. Power analyses can be used to help make this decision by estimating the number of sites to be sampled and replicates needed to produce a statistically significant result. The U.S. Geological Survey provides an excellent website that can assist the resource manager in determining the number of sites and replicates needed to produce the desired analytical power for a particular sampling design: http://www.mp1-pwrc.usgs.gov/powcase/how.html.

Table 1: Comparison of Stratified Random, Targeted, and BACI Sampling Designs

Stratified Random	Targeted	BACI
Random selection of wetland systems from entire population within a region.	Targeted selection of wetlands based on problematic (wetland systems known to have problems) and reference wetlands.	Selection of wetlands based on a known impact.
This design requires minimal prior knowledge of wetlands within the sample population for stratification.	This design requires prior knowledge of wetlands within the sample population.	This design requires knowledge of a specific impact to be analyzed.
This design may require more resources (time and money) to randomly sample wetland classes, because more wetlands may need to be sampled.	This design utilizes fewer resources because only targeted systems are sampled.	This design utilizes fewer resources because only wetlands with known impacts and associated control systems are sampled.
System characterization for a class of wetlands is more statistically robust.	System characterization for a class of wetlands is less statistically robust, although characterization of a targeted wetland may be statistically robust.	Characterization of the investigated systems is statistically robust.
Rare wetlands may be under-represented or absent from the sampled wetlands.	This design may miss important wetland systems if they are not selected for the targeted investigation.	The information gained in this type of investigation is not transferable to wetland systems not included in the study.
This design is potentially best for regional characterization of wetland classes, especially water quality conditions are not known.	This design is potentially best for site-specific and watershed-specific criteria development when water quality conditions for the wetland of interest are known.	This design is potentially best for monitoring restoration or creation of wetlands and systems that have specific known stressors.

Stratified Random Sampling Design

Probabilistic sampling—a sampling process in which randomness is a requisite (Hayek 1994)—can be used to characterize the status of water quality conditions and biotic integrity in a region's wetland systems. This type of sampling design is used to describe the average conditions of a wetland population, identify the variability among sampled wetlands, and to help determine the range of wetland system conditions in a region. However, the data collected from a probabilistic random sample design will generally be characteristic of the dominant class of wetland in the region, and rare wetlands may be underrepresented or absent from the probabilistically sampled wetlands. Additional sampling sites may need to be added to include the complete range of wetland conditions and classes in the region.

Probabilistic designs are often modified by stratification (such as classification). Stratification, or stratified random sampling, is a type of probability sampling in which a target population is divided into relatively homogenous groups (strata) before sampling based on factors that influence variability in that population (Hayek 1994). Analysis of variance can be used to identify statistically different parameter means among the sampling strata. The strata are the analysis of variance treatments (Poole

1972). The result of collecting and assessing water quality and biotic responses with a stratified random sample is, presumably, an unbiased estimate of the descriptive statistics (e.g., means, variances, modes, and quartiles) of all wetlands in a stratum. Stratification by wetland size and class provides more information about different types of wetlands within a region. Sample statistics from random selection alone would be most characteristic of the dominant wetland class in a region if the population of wetlands is not stratified.

Many State 305b and watershed monitoring programs use stratified random sampling designs, for example, Maine, Montana, and Wisconsin pilot projects use this design. Details of these monitoring designs can be found in the Module 14: Case Studies (Bioassessment) and at http://www.epa.gov/owow/wetlands/bawwg/index.html. Stratification is based on identifying wetland systems in a region (or watershed) and then selecting an appropriate sample of systems from the defined population. The determination of an appropriate sample population depends on the management questions being asked. A sample population of isolated depressional wetlands could be identified as a single stratum, but investigations of these wetlands would not provide any information on riparian wetlands in the same region. If the goal of the monitoring program is to identify wetland condition for all wetland classes within a region, then a sample population of wetlands should be randomly selected from all wetlands within each class. In practice, most State and Tribal programs stratify random populations by size, wetland class (see Module 7: Wetlands Classification), and landscape characteristics or location (see http://www.epa.gov/owow/wetlands/bawwg/case/me.html, http://www.epa.gov/owow/wetlands/bawwg/case/wa.html, and http://www.epa.gov/owow/wetlands/bawwg/case/wi1.html).

Once the wetlands for each stratum have been selected, the sample population is often modified by deleting systems that are too close to other wetlands to be different, thereby reducing redundant collection efforts. For example, the Environmental Monitoring and Assessment Program limits redundant collection efforts by applying a regular (hexagonal) grid to a map of the area. Sampling sites are chosen by randomly selecting grid cells and randomly sampling wetland resources within the chosen grid cells (Paulsen et al. 1991). Estimates of ecological conditions from these kinds of modified probabilistic sampling designs can be used to characterize the water quality conditions and biological integrity of wetland systems in a region and, over time, to distinguish trends in ecological condition within a region (see http://www.epa.gov/owow/wetlands/bawwg/case/mtdev.html and http://www.epa.gov/owow/wetlands/bawwg/case/fl1.html).

Targeted Design

A targeted approach to sampling design may be more appropriate when resources are limited. Targeted sampling is a specialized case of random stratified sampling. The approach described here involves defining a gradient of impairment. Once the gradient has been defined and systems have been placed in categories of impairment, investigators focus the most effort on identifying and characterizing wetland systems or sites likely to be impacted by anthropogenic stressors and on relatively undisturbed wetland systems or sites (see "Identifying and Characterizing Reference Wetlands") that can serve as regional, subregional, or watershed examples of natural biological integrity. The Florida Department of Environmental Protection uses a targeted sampling design for developing thresholds of impairment with macroinvertebrates (http://www.epa.gov/owow/wetlands/bawwg/case/fl2.html). Choosing sampling stations that best allow the comparison of ecological integrity at reference wetland sites of known condition can conserve financial resources. A sampling design that tests specific hypotheses (e.g., the study by the Florida

Department of Environmental Protection tested the effect of elevated water column phosphorus on macroinvertebrate species richness) can generally be analyzed with statistical rigor and can conserve resources by answering specific questions. Furthermore, the identification of systems with problems and reference conditions eliminates the need for selecting a random sample of the population for monitoring.

Targeted sampling assumes some knowledge of the systems sampled. Systems with evidence of degradation are compared with reference systems that are similar in physical structure (i.e., in the same class of wetlands). Targeted sampling requires that the wetlands be characterized by a gradient of impairment. Wetland systems should be placed along a continuum from reference to most impacted. An impaired or degraded wetland is simply a system in which anthropogenic impacts exceed acceptable levels or interfere with beneficial uses. Comparison of the monitoring data with the data collected from reference wetlands will allow characterization of the sampled systems. Wetlands identified as "at risk" should be evaluated through a sampling program to characterize the degree of degradation. Once characterized, the wetlands should be placed in categories such as the following:

- Degraded wetlands—wetlands in which the level of anthropogenic perturbance interferes with designated uses
- High-risk wetlands—wetlands in which anthropogenic stress is high but does not significantly impair designated uses (In high-risk systems, impairment is prevented by one or a few factors that could be changed by human actions, although characteristics of ecological integrity are already marginal.)
- Low-risk wetlands—wetlands in which many factors prevent impairment, stressors are maintained below problem levels, and/or no development is contemplated that would change these conditions
- Reference wetlands—wetlands in which the ecological characteristics most closely represent the pristine or minimally impaired condition.

Once wetland systems have been classified on the basis of their physical structure (see Module 7: Wetlands Classification) and placed into the categories previously defined, specific wetlands need to be selected for monitoring. At this point, randomness is introduced; wetlands should be randomly selected within each class and risk category for monitoring. An excellent example of categorizing wetlands in this manner is given in the Ohio Environmental Protection Agency's case study at http://www.epa.gov/owow/wetlands/bawwg/case/oh1.html. It used the Ohio Rapid Assessment Method to categorize wetlands by degree of impairment. The Minnesota Pollution Control Agency also used a targeted design for monitoring wetlands (see http://www.epa.gov/owow/wetlands/bawwg/case/mn1.html). It used the best professional judgment of local resource managers to identify reference sites as well as sites with known impairment from identified stressors (e.g., agriculture and stormwater runoff).

Monitoring efforts are often prioritized to best utilize limited resources. For example, case study investigators in Oregon chose not to monitor depressional wetlands because of funding constraints; they further tested the degree of independence of selected sites (and thus the need to monitor all of those sites) by using cluster analysis and other statistical tests (see http://www.epa.gov/owow/wetlands/bawwg/case/or.html). Frequency of monitoring is determined by the management question being asked and the intensity of monitoring necessary to collect enough information to answer the question. In addition, monitoring should identify the watershed-level activities that are likely to result in ecological degradation of wetland systems (Suter et al. 1993). Targeted sampling design involves monitoring identified degraded systems and comparable reference systems most intensively. Low-risk systems are monitored less frequently (after initial identification), unless changes in the watershed indicate an increased risk of degradation.

Activities surrounding impaired wetland systems may be used to help identify which actions negatively affect wetlands, and therefore may initiate more intensive monitoring of at-risk wetlands. Monitoring should focus on factors likely to identify ecological degradation and anthropogenic stress and on any actions that might alter those factors. State/Tribal water quality agencies should encourage adoption of local watershed protection plans to minimize ecological degradation of natural wetland systems. Development plans in a watershed should be evaluated to identify potential future stressors. Changes in point sources can be monitored through the National Pollutant Discharge Elimination System permit program (U.S. EPA 2000). Changes in nonpoint sources can be evaluated through the identification and tracking of wetland loss and/or degradation, increased residential development, increased tree harvesting, and shifts to more intensive agriculture with greater fertilizer use or increases in livestock numbers. Local planning agencies should be informed of the risk of increased anthropogenic stress and encouraged to guide development accordingly. Ecological degradation often gradually increases as a result of many growing sources of anthropogenic stress. Therefore, frequent monitoring is warranted for high-risk wetlands if sufficient resources remain after meeting the needs of degraded wetlands. Whenever development plans appear likely to alter factors that maintain ecological integrity in a high-risk wetland (e.g., vegetated buffer zones), monitoring should be initiated at a higher sampling frequency in order to enhance the understanding of baseline conditions (U.S. EPA 2000).

BACI Survey Design

An ideal impact survey has several features: The type of impact, time of impact, and place of occurrence should be known in advance; the impact should not have occurred yet; and control areas should be available (Green 1979). The first feature allows surveys to be efficiently planned to account for the probable change in the environment. The second feature allows a baseline study to be established and to be extended as needed. The third feature allows the surveyor to distinguish between temporal effects unrelated to the impact and changes related to the impact. In practice, however, advance knowledge of specific impacts is rare and the ideal impact survey is rarely conducted. BACI designs modified to monitor impacts during or after their occurrence can still provide information, but there is an increase in the uncertainty associated with the results, and the likelihood of finding a statistically significant change caused by the impact is much less probable. Power analyses of after-only studies were conducted by Osenberg et al. (1994). They determined that because of the time constraints of most studies, relatively few of the population and chemical/physical parameters could provide adequate analytical power. They suggest expending a greater effort on monitoring individual-based parameters (e.g., body size and recruitment density) in addition to population and chemical/physical parameters for environmental impact assessments (Osenberg et al. 1994). Defining the study objectives, and identifying the specific hypotheses being tested, greatly increases the certainty of the results. In addition, other aspects of survey design are dependent on the study objectives: the sampling interval, the length of time the survey is conducted (i.e., sampling for acute vs. chronic effects), and the statistical analyses appropriate for analyzing the data (Suter 1993).

The best interval for sampling is determined by the objectives of the study (Kentula et al. 1993). If the objective is to detect changes in trends (e.g., regular monitoring for detection of changes in water quality or biotic integrity), regularly spaced intervals are preferred because the analysis is easier. However, if the objective is to assess differences before and after impact, then samples at random time points are advantageous. Random sample intervals reduce the likelihood that cyclic differences unforeseen by the sampler will influence the size of the difference before and after the impact. For example, surveys taken every summer for several years before and after a clear-cut may show little difference in system quality; however, differences may exist that can be detected only in the winter and, therefore, they may go undetected if sampling occurs only during summer.

The simplest impact survey design involves taking a single survey before and after the impact event (Green 1979). This type of design has the obvious pitfall that there may be no relationship between the observed event and the changes in the response variable—the change may be entirely coincidental. This pitfall is addressed in BACI design by comparing before and after impact data with data collected from a similar control system nearby. Data are collected before and after a potential disturbance in two areas (treatment and a control), with measurements on biological and environmental variables in all combinations of time and area (Green 1979). For example, consider a study in which the investigators want to identify the effects of clear-cutting on wetland systems. In the simplest BACI design, two wetlands would be sampled. One wetland would be adjacent to the clear-cut (the treatment wetland); the other wetland would be adjacent to a control site that is not clear-cut. The control site should have characteristics (i.e., soil, vegetation, structure, and functions) similar to the treatment wetland and should be exposed to climate and weather similar to the first wetland. Both wetlands are sampled at the same time points before and after the clear-cut occurs. This design is technically known as an area-by-time factorial design. Evidence of an impact is found by comparing the control site samples (before and after) with the treatment site before and after samples. Area-by-time factorial design allows for both natural wetland-to-wetland variation and coincidental time effects. If the clear-cut has had no effect, then the change in system quality between the two time points should be the same. If the clear-cut has had an effect, then the change in system quality between the two time points should be different.

There are some potential problems with BACI design. First, because the control and impact sites are not randomly assigned, observed differences between sites may be related solely to some other factor that differs between the two sites. One could argue that it is unfair to ascribe the effect to the impact (Hurlbert 1984, Underwood 1991). However, as pointed out by Stewart-Oaten et al. (1986), the survey is concerned about a particular impact in a particular place, not about the average of several impacts when the survey is replicated in many different locations. Consequently, it may be possible to detect a difference between these two specific sites. However, if there are no randomized replicate treatments, the results of the study cannot be generalized to similar events at different wetlands. However, the likelihood that the differences between sites are due to factors other than the impact can be reduced by monitoring several control sites (Underwood 1991). If one assumes that the variation in the (before and after) measurements of multiple control sites is the same as the variation among potentially impacted sites, and that the variability over time between the control sites is not correlated, one can estimate the likelihood that the impact caused the observed difference at the impacted site, given the observed variability in the control sites. That is, several control wetlands could be monitored at the same time points as the single impact wetland. If the observed difference in the impact wetland is much different than could be expected based on the multiple control wetlands, the event is said to have caused an impact. The lack of randomization is less of a concern when several control sites are monitored, because the multiple control sites provide some information about potential effects of other factors.

The second and more serious concern with the simple before and after design with a single sampling point before and after the impact is that it fails to recognize that natural fluctuations may occur in the characteristic of interest that are unrelated to any impact (Hurlbert 1984, Stewart-Oaten 1986). Single samples before and after impact would be sufficient to detect the effects of the impact if no natural fluctuations occurred over time. However, if the population also has natural fluctuations over and above the long-term average, then it is impossible to distinguish between cases in which no effect occurs from cases in which an impact does occur. Consequently, measured differences in system quality may be artifacts of the sampling dates, and natural fluctuations may obscure differences or lead one to believe differences are present when they are not.

The simple BACI design was extended by Stewart-Oaten et al. (1986) by pairing surveys at several selected time points before and after the impact to help resolve the issue of pseudoreplication (Hurlbert 1984). This modification of the BACI design is referred to as the BACI-paired series (PS) design. The selected sites are measured at the same time points. The rationale behind this paired design is that repeated sampling before the impact gives an indication of the pattern of differences of potential change between the two sites. BACI-PS study design provides information on the mean difference in the wetland system quality before and after impact and on the natural variability of the system quality measurements. An effect is detected if the changes in the mean difference are large relative to natural variability. Considerations for sampling at either random and regularly spaced intervals also apply here.

BACI-PS study design also has potential pitfalls. As with all studies, numerous assumptions need to be made during the analysis (Stewart-Oaten et al. 1992, Smith et al. 1993). The primary assumption for BACI-PS design is that the responses over time are independent of each other. A lack of independence over time tends to produce false-positive (Type I) errors, which may lead a manager to declare that an effect has occurred when, in fact, none has. Formal time series analysis methods or repeated measures analysis may be necessary (Rasmussen et al. 1993) to eliminate Type I errors. (The analysis of time series is easiest with regularly spaced sampling points.) In addition, the difference in mean level between control and impact sites is assumed to be constant over time in the absence of an impact effect. The effect of the impact is assumed to change the arithmetic difference. In the clear-cut example given previously, the difference in mean system quality between the two sites is assumed to be constant over time. That is, mean system quality measurements may fluctuate over time, but both sites are assumed to fluctuate in the same manner simultaneously, thereby maintaining the same average arithmetic difference. This assumption is violated if the response variable at the control site is a constant multiple of the response variable at the impact site. For example, suppose that the readings of water quality at two sites at the first time point were 200 vs. 100, which has an arithmetic difference of 100, and at the second time point were 20 vs. 10, which has an arithmetic difference of 10, but both pairs are in a 2:1 ratio at both time points. The remedy is simple: A logarithmic transform of the raw data converts a multiplicative difference into a constant arithmetic difference on the logarithmic scale. This is a common problem when system quality measurements are concentrations (e.g., pH). Smith et al. (1993) pointed out that this may not solve the issue of pseudoreplication. Trends are common in most natural populations, but BACI design assumes that trends are not present in the populations or that the control and impact sites have the same trends, so that differences between the sites are identified as associated with the impact, not with differences in trends of natural populations (Smith et al. 1993). Violation of the BACI assumptions may invalidate conclusions drawn from the data. Enough data must be collected before the impact to identify the trends in the communities of each sampling site if the BACI assumptions are to be met. Clearly defining the objectives of the study and identifying a statistically testable model of the relationships the investigator is studying can help resolve these issues (Suter 1993).

Underwood (1991) also considered two variations on the BACI-PS design. First, it may not be possible to sample both sites simultaneously for technical or logistical reasons. Underwood (1991) discussed a modification in which sampling is done at different times in each site before and after impact (i.e., sampling times are no longer paired), but notes that this modification cannot detect changes that occurred in the two sites before the impact. For example, differences in system quality may show a gradual change over time in the paired design before impact. Without paired sampling, it would be difficult to detect this change. In addition, sampling only a single control site still has the problem identified previously, that is, it is not known whether observed differences in the impact and the control sites are site specific. Again, Underwood (1991) suggests that multiple control sites should be monitored. The variability in the difference between each control site and the impact site provides information on transferability of the impact effects to other sites (i.e., it either refutes or supports the site specificity of the impact and associated system response).

The designs described are suitable for detecting long-term, or chronic, effects in the mean level of the variable of interest. However, the impact may have a short-term, or acute, effect, or it may change the variability in response (e.g., seasonal changes become more pronounced) in some cases. The sampling schedule can be modified to occur at two temporal scales (enhanced BACI-PS design) that encompass both acute and chronic effects (Underwood 1991). The modified temporal design introduces randomization by randomly choosing sampling occasions in two periods (before and after) in the control or impacted sites. The two temporal scales (sampling periods vs. sampling occasions) allow the detection of a change in mean and in variability after impact. For example, groups of surveys could be conducted every year, with five surveys one week apart randomly located within each group. The analysis of such a design is presented in Underwood (1991). Again, multiple control sites should be used to confound the argument that detected differences are specific to the sampled site.

BACI-PS design is also useful when there are multiple objectives. For example, the objective for one variable may be to detect a change in trend. The pairing of sample points on a long time scale leads to efficient detection of trend changes. The objectives for another variable may be to detect differences in the mean level. A short time scale surveys randomly located in time and space are efficient for detecting differences in the mean level. The September 2000 issue of the *Journal of Agricultural, Biological, and Environmental Statistics* discusses many of the advantages and disadvantages of the BACI design and provides several examples of appropriate statistical analyses for evaluation of BACI studies.

Suggested Websites

1 http://ebook.stat.ucla.edu/calculators/powercalc/

2 http://www.math.sfu.ca/stats/Courses/Stat-650/Notes/Handouts/node1.html

3 http://www.mp1-pwrc.usgs.gov/powcase/index.html

4 http://www.salmonweb.org/salmonweb/pubs/pacnwfin.html

5 http://trochim.human.cornell.edu/tutorial/flynn/multivar.htm

6 http://www.tufts.edu/~gdallal/STUDY.HTM

7 http://www.umass.edu/tei/mwwp/studydes.html

References

Adamus PR. 1992. Choices in Monitoring Wetlands. In: McKenzie DH, Hyatt DE, McDonald VJ (eds). Ecological Indicators. New York: Elsevier Applied Science, pp. 571-592.

Barbour MT, Gerritsen J, Snyder BD (eds). 1999. Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, 2nd ed. U.S. Environmental Protection Agency, Washington, DC. EPA 841-B-99-002.

Beyers DW. 1998. Causal inference in environmental impact studies. J North Am Benthol Soc 17:367-373.

Davis SM, Ogden JC (eds). 1994. Everglades: The Ecosystem and Its Restoration. Delray Beach, FL: St. Lucie Press.

Detenbeck NE, Taylor DL, Lima A. 1996. Spatial and temporal variability in wetland water quality in the Minneapolis/St. Paul, MN, metropolitan area. Environ Monitor Assess 40:11-40.

Dodds WK, Smith VH, Zander B. 1997. Developing nutrient targets to control benthic chlorophyll levels in streams: a case study of the Clark Fork River. Water Res 31:1738-1750.

Eaton AD, Clesceri LC, Greenberg AE (eds). 2000. Standard Methods for Examination of Water and Wastewater, 19th ed. Washington, DC: American Public Health Association.

Green RH. 1979. Sampling Design and Statistical Methods for Environmental Biologists. New York: Wiley.

Gwin SE, Kentula ME, Shaffer PW. 1999. Evaluating the effects of wetland regulation through hydrogeomorphic classification and landscape profiles. Wetlands 19(3):477-489.

Harris SC, Martin TH, Cummins KW. 1995. A model for aquatic invertebrate response to Kissimmee River restoration. Restoration Ecol 3:181-194.

Hayek LC. 1994. Research Design for Quantitative Amphibian Studies. In: Measuring and Monitoring Biological Diversity, Standard Methods for Amphibians. Washington, DC: Smithsonian Institution Press.

Hurlbert SH. 1984. Pseudo-replication and the design of ecological field experiments. Ecol Monogr 52:187-211.

Johnston CA, Detenbeck NE, Bonde JP, Niemi GJ. 1988. Geographic information systems for cumulative impact assessment. Photogr Engin Remote Sens 54:1609-1615.

Johnston CA, Detenbeck NE, Niemi GJ. 1990. The cumulative effect of wetlands on stream water quality and quantity: A landscape approach. Biogeochemistry 10:105-141.

Kadlec RH, Knight RL. 1996. Treatment Wetlands. Boca Raton, FL: Lewis Publishers.

Karr JR, Chu EW. 1999. Restoring Life in Running Waters: Better Biological Monitoring. Washington, DC: Island Press.

Kentula ME, Brooks RP, Gwin SE, Holland CC, Sherman AD, Sifneos JC. 1993. An Approach to Improving Decision Making in Wetland Restoration and Creation. Boca Raton, FL: CK Smoley.

Leibowitz SG, Abbruzzese A, Adamus PR, Hughes LE, Irish JT. 1992. A Synoptic Approach to Cumulative Impact Assessment: A Proposed Methodology. Environmental Research Laboratory, U.S. Environmental Protection Agency, Corvallis, OR. EPA/600/R-92/167.

McCormick P, Newman S, Miao S, Reddy R, Gawlick D, Fitz C, Fontaine T, Marley D. 1999. Ecological Needs of the Everglades. In: Redfield G (ed). Everglades Interim Report. South Florida Water Management District, West Palm Beach, FL.

Mitsch WJ, Gosselink JG. 1993. Wetlands, 2nd ed. New York: Van Nostrand Reinhold.

Osenberg CW, Schmitt RJ, Holbrook SJ, Abu-Saba KE, Flegal R. 1994. Detection of environmental impacts natural variability, effect size, power analysis. Ecol Appl 4:16-30.

Palik AJ, Goebel CP, Kirkman KL, West L. 2000. Using landscape hierarchies to guide restoration of disturbed ecosystems. Ecol Appl 10(1):189-202.

Paulsen SG, Larsen DP, Kaufmann PR, Whittier TR, Baker JR, Peck DV, McGue J, Hughes RM, McMullen D, Stevens D, Stoddard JL, Lazorchak J, Kinney W, Selle AR, Hjort R. 1991. EMAP-Surface Waters Monitoring and Research Strategy: Fiscal year 1991. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC. EPA/600/3-91/022.

Poole RW. 1972. An Introduction to Quantitative Ecology. New York: McGraw-Hill.

Rasmussen PW, Heisey DM, Nordheim EV, Frost TM. 1993. Time Series Intervention Analysis: Unreplicated Large-scale Experiments. In: Scheiner SM, Gurevitch J (eds). Design and Analysis of Ecological Experiments. New York: Chapman and Hall.

Redfield G (ed). 1999. Everglades Interim Report. South Florida Water Management District, West Palm Beach, FL.

Redfield G (ed). 2000. Everglades Consolidated Report. South Florida Water Management District, West Palm Beach, FL.

Redfield G (ed). 2001. Everglades Consolidated Report. South Florida Water Management District, West Palm Beach, FL.

Smith EP, Orvos D, Cairns J, Jr. 1993. Impact assessment using before-after control-impact (BACI) models: Concerns and comments. Can J Fish Aquat Sci 50:627-637.

Stevenson RJ. 1996. An Introduction to Algal Ecology in Freshwater Benthic Habitats. In: Stevenson RJ, Bothwell M, Lowe RL (eds). Algal Ecology: Freshwater Benthic Ecosystems. San Diego, CA: Academic Press, pp. 3-33.

Stevenson RJ. 1997. Scale dependent determinants and consequences of benthic algal heterogeneity. J North Am Benthol Soc 16(1):248-262.

Stewart-Oaten A. 1986. Assessing local impacts: progress and some problems. Oceans ’86 Conference Record 3:964-973.

Stewart-Oaten A, Bence JR, Osenberg CW. 1992. Assessing effects of unreplicated perturbations: No simple solutions. Ecology 73:1396-1404.

Stewart-Oaten A, Murdoch WW, Parker KR. 1986. Environmental impact assessment: “Pseudoreplication” in time? Ecology 67:929-940.

Suter GW. 1993. Ecological Risk Assessment. Boca Raton, FL: Lewis Publishers.

Tate CM. 1990. Patterns and controls of nitrogen in tallgrass prairie streams. Ecology 71:2007-2018.

Toth LA, Arrington DA, Brady MA, Muszick DA. 1995. Conceptual evaluation of factors potentially affecting restoration of habitat structure within the channelized Kissimmee River ecosystem. Restoration Ecol 3:160-180.

Trexler JC. 1995. Restoration of the Kissimmee River—A conceptual model of past and present fish communities and its consequences for evaluating restoration success. Restoration Ecol 3:195-210.

Underwood AJ. 1991. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. Austral J Marine Freshw Res 42:569-587.

Urquhart NS, Paulsen SG, Larsen DP. 1998. Monitoring for policy-relevant regional trends over time. Ecol Appl 8(2):246-257.

U.S. Environmental Protection Agency (EPA). 1990. Water Quality Standards for Wetlands: National Guidance. Office of Water, Washington, DC. EPA 440/5-90-011.

U.S. EPA. 1996. National Water Quality Inventory 1996 Report to Congress. Office of Water, Washington, DC. EPA 841-R-97-008.

U.S. EPA. 1998. National Water Quality Inventory 1998 Report to Congress. Office of Water, Washington, DC. EPA 841-S-00-001.

U.S. EPA. 2000. Nutrient Criteria Technical Guidance Manual: Rivers and Streams. Office of Water, Washington, DC. EPA 822-B-00-002.