Chapter 3. Planning and Design for Habitat Monitoring

Christina D. Vojta
Lyman L. McDonald
C. Kenneth Brewer
Kevin S. McKelvey
Mary M. Rowland
Michael I. Goldstein

3.1 Objective

This chapter provides guidance for designing a habitat monitoring program so that it will meet the monitoring objective, will be repeatable, and will adequately represent habitat within the spatial extent of interest. Although a number of excellent resources are available for planning and designing a monitoring program for wildlife populations (e.g., Busch and Trexler 2003, McComb et al. 2010, Thompson et al. 1998, Vesely et al. 2006), little guidance exists for creating a monitoring design for habitat. One could argue that the huge body of literature on vegetation sampling is useful for this purpose, and we acknowledge that the basic principles of planning and design for vegetation sampling are indeed relevant. Most texts on vegetation sampling do not address the multiscalar nature of habitat, however, or the nonvegetative aspects of habitat. Moreover, unique challenges arise related to using existing data and to using remotely sensed data for monitoring habitat. Our objective is to emphasize the basic principles of planning and design and to place them in the context of habitat monitoring.

In all aspects of planning and design, we recommend early consultation with a statistician. Too often, monitoring teams consult with a statistician only after data are collected. Seemingly innocent decisions in the design or placement of plots or points, however, can make statistical analyses exceptionally complex or impossible. Early consultation will ensure that statistical analyses are feasible and practical before implementing the monitoring design.

3.2 Key Concepts

3.2.1 Central Coordination and Long-Term Commitment

Long-term commitment and central coordination are paramount to a successful monitoring program, especially long-term programs. An agency or organization needs to have ownership in the program and an individual or team needs to ensure that the program is carried out over time and that it will persist through funding obstacles. Successful programs often are driven by someone who has personal vision and commitment, but motivation can also come from an entity that is funded specifically for the monitoring purpose, such as the Breeding Bird Survey (BBS) that is coordinated annually by the U.S. Geological Survey (USGS) and the Canadian Wildlife Service (Sauer et al. 2008) or the Forest Inventory and Analysis (FIA) program funded by the Forest Service and numerous partners (chapter 4, section 4.4.1). Sometimes a combination of personal vision and consistent funding leads to success. For example, the 20-year success of the Northern Region Landbird Monitoring Program, a cooperative effort between the Forest Service and the University of Montana, was the result of a combination of individual leadership coupled with adequate funding for planning, training, data collection, and data analysis.

We begin this chapter on planning and design with this key concept, because it is during the planning and design phases that central coordination and long-term commitment must be identified. Long-term funding is rarely assured, but long-term commitment and ownership from individuals and organizations are essential.

3.2.2 Sources of Error and Uncertainty in Habitat Monitoring

A well-designed monitoring program will consider possible sources of error and uncertainty and reduce them where possible. The first source of uncertainty arises during the development of the species' conceptual model and the selection of habitat attributes to monitor (chapter 2). This form of uncertainty arises because population dynamics can never be modeled with absolute precision, and measured habitat attributes will always be a subset of the species' total habitat requirements. It is not possible to reduce this uncertainty through a sampling design. The best approach is to evaluate the underlying assumptions of what constitutes habitat by surveying for species' presence across a gradient of environmental conditions (chapter 1 and 5).

Another source of error occurs when habitat attributes are not directly measured but are derived through modeling. This topic is sufficiently complex to warrant treatment as a separate key concept (3.2.6).

In this section, we address two sources of error that can be somewhat managed through sampling design and data collection—environmental variability (also known as **process variation)** and **sampling variation (**White 2000). Biometricians originally coined the term *process variability* to explain the combined effects of demographic, temporal, spatial, and individual variation on wildlife populations (White 2000). In the scope of habitat monitoring, process variability is encountered when repeated estimates occur at different times. Temporal changes are expected to occur in the true values of the attributes regardless of the direction of long-term trends over a large area. For example, suppose we have three standing snags in a plot. Between time one and time two, the number of standing snags within this plot may change, regardless of whether the long-term trend is increasing, constant, or decreasing in the larger study area. Long-term trends may be of interest, whereas annual fluctuations and other sources of process variation might be considered a nuisance. If the objective is a habitat inventory rather than monitoring, the estimates apply to a fixed study area and, ideally, to an instantaneous moment in time. Process variability still introduces uncertainty, however, in the sense that a single estimate at a point in time will fall below, on, or above the unobserved long-term trend line or curve.

Sampling variability occurs during data collection and analysis when one is estimating the true spatial distribution, abundance, or density of each habitat attribute from a sample at a specific point in time (in practice, a period during which the attributes are assumed to not change appreciably). This uncertainty stems from two sources: **variance** (i.e., precision; how close repeated measurements are to the same value) and **bias** (inversely related to accuracy, which is how close the average value of measurements is to the actual value of an attribute) (Thompson et al. 1998, Zar 2010). The variance of an estimate can also be attributed to two sources: environmental variance that stems from sampling the study area (because the true values of an attribute vary from one sampling unit to another) and measurement error (because any two independent measurements may yield different numerical values).

Bias stems from flaws in the sample design or measurement protocol that result in an estimate either consistently higher or lower than the true value. For example, the sample design can result in bias if sampling units are not randomly selected or represent only part of the area of interest (see Spatial Extent in the following section). The measurement protocol may also have inherent positive or negative bias relative to the true value (for instance, when an instrument is not calibrated correctly) (Elzinga et al. 1998).

Improving the measurement protocol can reduce bias and measurement errors. For example, bias in the measurement of forb abundance might be reduced by considering the phenology of the plants and by ensuring that all resampling occurs at the same phenological stage. Including more units from the study area can reduce sampling error. Sampling error, measurement error, and process variability are usually confounded, however, and cannot be separated without additional experimentation. Consequently, in most simple modeling of long-term trends, we crudely incorporate uncertainty by estimating the magnitude of variation arising from a combination of factors—process variability, sampling variance, and measurement error. Process variability can sometimes be separated from sampling variance and measurement error in relatively rich data situations using rather strict assumptions (e.g., Bolker 2008); however, those models and methods are beyond the scope of this text.

3.2.3 Competing Designs for Inventory and Monitoring

Although natural resource managers often view monitoring as a series of inventories across time, the optimal statistical design for assessing the status of a resource (inventory) can be very different from the design for detecting change (monitoring). Inventory design focuses on obtaining a representative sample of a **sampling frame** and controlling sampling variance and bias through randomization and the selection of independent samples. In contrast, monitoring can be directed at specific areas with a different sampling frame for the purpose of trend detection without providing an overall estimate of the current status of a resource.

Consider, for example, the loss of meadows to conifer encroachment, which is a common problem for land managers. For an inventory, the question would be: What is the current area in meadows? For monitoring, the question might be: How rapidly is the meadow area changing? One could, of course, evaluate change in meadow area across time through multiple inventories, by randomly sampling from a sampling frame that contains all potential meadow area, but a much more efficient and sensitive approach to detect change would be to place plots at the meadow edges—the only place on the landscape where change is actually occurring, and revisit those plots. A different sampling frame of all units containing meadow edge would be constructed and sampling units selected by a probabilistic (e.g., simple random) sampling procedure. Note that this second approach will not provide an estimate of the total acreage of meadows in the area of interest.

3.2.4 Independent Samples Versus Repeated Measures

A fundamental choice in design of monitoring studies is between measurements on independent samples of units at each time point and repeated measurements on the same units (for two or more surveys over time). We will illustrate this choice using snag density as an example. If different independent samples of units are selected at each point in time, then the monitoring question of whether the number of snags per acre has changed between two points in time requires having two independent estimates, one for each point in time. Each of the estimates has an associated variance, as indicated with the subscripts 1 and 2 in the following equation for two independent samples. For moderate sample sizes, an approximate **confidence interval** on the difference is

$$(\overline{X}_1 - \overline{X}_2) \pm z_{0.25}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\ .$$

The variances are additive for the difference in means of independent samples and, if the variances associated with the two estimates are approximately the same (as expected to be if the same sampling design were used for both samples), the variance associated with a difference will be about double the variance of either point estimate for the number of snags per acre. This larger variance means that, with independent samples, it can be difficult to detect a meaningful change in the number of snags per acre (or any other habitat attribute) between two points in time.

A further difficulty with selecting independent samples at each time point is that, when attempting to detect changes in a habitat attribute across relatively short periods of time (e.g., 5 to 15 years for forest plan monitoring), the difference between two sample means $(\overline{X}_1 - \overline{X}_2)$ will usually be relatively small. Thus, detecting change across short time intervals with independent samples may be difficult, because the difference often will be a number relatively close to 0.

In contrast, if repeated measurements are obtained on the same sampled units, then the data are essentially paired, and the analysis for detecting change between two points in time is conducted on the differences, $D_i = X_{i1} - X_{i2}$, for all units (plots) i = 1, …, n. The mean difference, $\overline{D} = \overline{X}_1 - \overline{X}_2$, and the variance s_D^2 of the differences are computed. A similar formula, but for a single sample of differences, unit by unit, is used to compute a confidence interval on the mean difference, i.e., $\overline{D} \pm z_{0.25}\sqrt{\frac{s_D^2}{n}}$.

With repeated measurements on the same sampled units, the variance, s_D^2, and, consequently, the **standard error** of the mean, $s_D/\sqrt{n}$, are usually much smaller for the differences, $D_i = X_{i1} - X_{i2}$, than the corresponding variance and standard error for either measures X_{i1} *or* X_{i2} at the two points in time. Consequently, as can be seen with the example of snag density, the width of the confidence interval on the mean difference is expected to be less than the widths of confidence intervals on the individual mean number of snags per acre at either time point. More importantly, the width of the confidence interval on the mean difference of any habitat attribute metric is usually much less with repeated measurements than for measurements on independent samples.

In conclusion, repeated measurements on the same units have the advantage over use of independent samples if it is important to detect change and trend more quickly in monitoring studies. The choice is not always easy, however, because measurements on independent samples have the advantage if it is more important to have better coverage of units throughout a study area over time. In the design of monitoring programs, usually the need to quickly detect change and trend is considered to be more important than ensuring independence of sampling units because of pressing management issues.

3.2.5 Monitoring for Thresholds Versus Trends

Although the customary objective of monitoring is to detect a trend, an alternative is to monitor for the relationship between a current condition and a preestablished **threshold** value in one or more habitat attributes to evaluate whether a change in management is needed. The first objective typically requires relatively long-term monitoring, whereas the second objective can be achieved with an inventory or at any point within a short- or long-term monitoring program.

To illustrate the difference, consider two different monitoring objectives regarding American marten (*Martes americana*) habitat. The question of whether the amount of habitat has changed requires the ability to detect change over time and may necessitate monitoring over several years or decades unless the change is abrupt (e.g., wildfire). In contrast, the question of whether 50 percent of forested land within an analysis area is suitable for American martens requires the ability to make a reasonable point estimate, which can then be compared with the 50-percent threshold or desired condition value at the end of one survey or at any point in time during a monitoring program.

The ability to detect changes in habitat over time becomes more tractable over many years, when samples are collected across many time points, allowing trend lines or curvilinear models to be fitted. The question then becomes whether the slope of a trend line is statistically different from zero; the answer is obtained through regression analysis.

3.2.6 Measured Versus Modeled Attributes

Habitat attributes for each emphasis species may be either measured directly or estimated through modeling. This distinction is critical because the source of data can affect the precision associated with an observed change in habitat over time. Measured attributes include original data and simple summary statistics of these data (e.g., sums, means, proportions, and variances).

Modeled attributes are derived from measured data by using statistical modeling processes or by assuming relationships of measured to unmeasured data through professional judgment (e.g., **Bayesian belief networks** in which prior distributions of parameters are assigned through a combination of expert opinion and existing data). A statistical modeling process estimates the correlation between two or more measured attributes and then uses this correlation to model or predict the value of an unmeasured attribute. For example, the correlation between the number of trees tallied with a basal area factor prism and the actual, measured basal area of the stand provides a modeled relationship that enables surveyors to obtain stand basal area from the number of trees tallied with a basal area factor prism.

Examples of commonly measured attributes are tree diameters, shrub height, and miles of maintained roads. Examples of modeled attributes are growth and yield based on measured tree heights and diameters, the amount and spatial distribution of vegetation types based on classification of **multispectral data,** and the amount and distribution of rainfall within a monitoring area based on rainfall measured at weather stations (point locations). The line between modeled and measured attributes is often blurry: although elevation can be a measured attribute, it can also be derived from radar through modeling.

Maps of classified vegetation are probably the most common source of modeled attributes of wildlife habitat. When using satellite imagery, the measured data are light reflectance values, whereas habitat classifications based on these light reflectance values are modeled data. The modeled data incur error based on a variety of factors: time of day and year, haziness, aspect, cloud cover, and other environmental conditions (Warbington 2011). For both satellite imagery and high-resolution aerial photography, additional error occurs when a map specialist classifies pixels as vegetation types because this form of classification is usually based on professional judgment and is often undefined in an absolute sense. As a result, small changes in the definition of a given vegetation type can have large effects on the amount of that vegetation type assigned to the map. This potential source of error explains why ground truthing is so important when developing maps of classified vegetation.

Because modeled attributes are derived rather than directly measured, the calculations and modeling assumptions add layers of imprecision to estimates of the attribute values. If the models are complex, it may be impossible to evaluate variance or standard errors of the modeled attributes. In addition, some sources of error can never be incorporated into measures of precision because the error is from professional judgment in selecting which form of model to use, the methods of estimating **coefficients**, and the process of setting criteria for selection among fitted models. The error generated from these decisions is less obvious than the estimated standard errors around the modeled attributes, but can have huge influence on the modeled attribute values and precision.

The problems associated with modeled attributes are compounded further when the attributes are used to monitor change over time. It is important to know whether the maps for each time period are derived from the same data sources, are based on the same statistical estimation methods, and use the same classification system for vegetation types or other attributes, because changes in any of these conditions can make it difficult to compare the values of modeled attributes over time. Modeled attributes are often preferred because they are less expensive to acquire than measured attributes. They should be used with caution, however, because of the many potential sources of error that we have described.

3.2.7 Cause-and-Effect Monitoring

As stated in chapter 1, cause-and-effect monitoring is outside the scope of this technical guide because it requires a fully developed experimental design (Elzinga et al. 1998, Holthausen et al. 2005). Noon (2003) described the concept of cause-and-effect monitoring and contrasted prospective and retrospective approaches to investigating cause and effect. Prospective approaches use experimental designs that assign treatment and reference sites to study units through a probabilistic method and are usually limited to laboratory studies and field studies with small plots. The design must incorporate sufficient replicates of each treatment as well as replicates of the reference or control areas to support any statistical conclusions and inferences on cause-and-effect relationships. Most individual national forests and grasslands will have insufficient replications of similar treatments or management actions to yield a high level of confidence in conclusions, but this type of monitoring can be accomplished over broad spatial scales with involvement from the research community. One example is the Birds and Burns project (http://www.rmrs.nau.edu/wildlife/birdsandburns/), which has established burn treatments and controls on several national forests and other lands in the Western United States (Saab et al. 2007a).

In contrast to cause-and-effect experimental designs, monitoring programs for detecting a trend or observing progress toward or away from a threshold are usually unreplicated observational studies. Sampling units are selected through a probabilistic or systematic procedure, resulting in units that represent the entire area of interest—managed and unmanaged. If the managed and unmanaged areas are similar except for the management action, strong conjectures are possible concerning the effectiveness of management. Unfortunately, the effects of management actions are nearly always confounded with inherent differences between sites in observational studies. Only with replication of treatments or the existence of other collaborating information will managers have the evidence necessary to convince critical peers that the results and strong conjectures have actually led to knowledge of cause-and-effect relationships concerning the management actions.

3.3 Process Steps for Designing a Habitat Monitoring Program

Use the following key steps to design a habitat monitoring program.

- Develop the monitoring objective.
- Evaluate whether existing information can be used to meet the monitoring objective.
- Plan and design for new data collection, where needed.
- Run a pilot study (or evaluate the statistical properties of existing data).
- Document the design decisions in a monitoring plan.

This section provides guidance for completing each of these steps. Figure 3.1 illustrates the process steps and some of the major decisions associated with the key concepts listed previously.

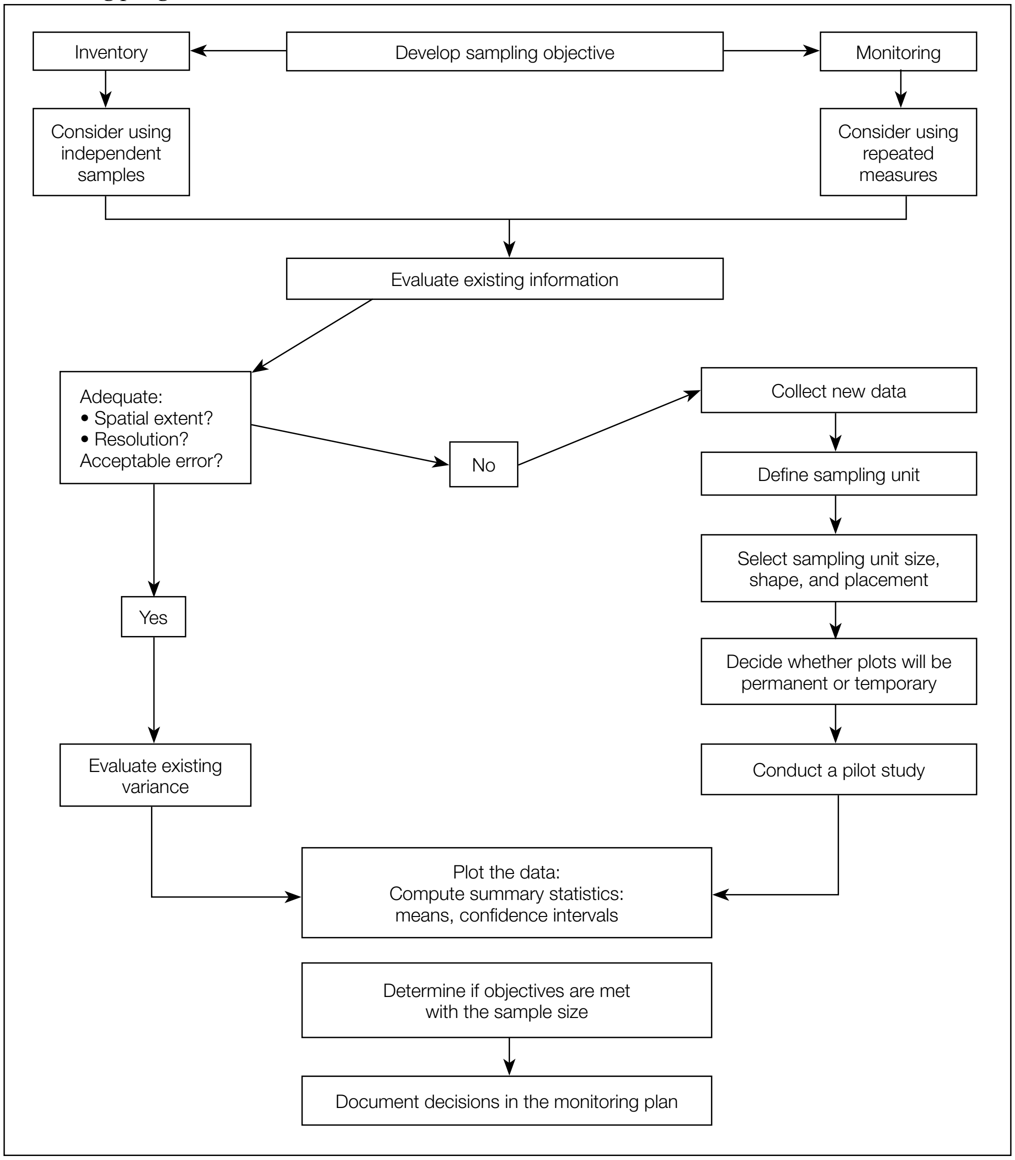
3.3.1 Develop the Monitoring Objective

After identifying the emphasis species and their key habitat attributes (chapter 2), the next task is to state the habitat monitoring objective with sufficient detail so that decisions can be made about how data will be collected and analyzed. A detailed objective is requisite for a well-designed habitat monitoring program and is critical to its success (Elzinga et al. 1998, Oakley et al. 2003). The monitoring objective provides a direct link between the management objective and the resource being monitored (e.g., a habitat attribute, such as large snag density). The following components comprise a monitoring objective (Elzinga et al. 1998, Vesely et al. 2006).

- The desired information outcome of the monitoring activity.
- The spatial extent over which the information is needed.
- The level of confidence or precision desired to fulfill management information needs.
- The desired **minimum detectable change** (or **effect size**); i.e., the size of change that is biologically meaningful.
- A quantitative standard against which the monitoring results would be compared in the short term; e.g., a desired condition, historical condition, or threshold that would trigger management actions in the near future.

Each of these objective components is described in the following sections.

Figure 3.1.—*Process steps and major decision points that occur during the development of a monitoring program.*



Desired Information Outcome

First, determine whether the monitoring program will be a one-time inventory or a multiyear effort. For a one-time inventory of a large study area, **stratification** can be a useful tool for reducing variance or for blocking on potential confounding factors, but this same tool can cause problems for multiyear monitoring programs, especially if strata boundaries are based on ephemeral conditions such as existing vegetation type. Differences between an inventory and a multiyear monitoring program can also influence the decision of marking plots. Multiyear monitoring may incur the added cost of marking permanent plots, with the benefit, however, that revisiting the same sample plots will usually detect important changes or trends more quickly than visiting new plots each sampling period.

Next, decide if individual habitat attributes should be tracked separately, or if it is more meaningful to aggregate all selected attributes in a model that defines habitat quality. This decision can influence whether the final product can simply be tabular data that summarizes the amount of each selected habitat attribute, or whether the data need to be mapped to evaluate spatial patterns.

If the monitoring objective is to observe a specific level of change, then sampling must be designed to determine whether the target has been met. If a plan states that a threshold must not be crossed (e.g., a maximum road density or minimum canopy cover), then monitoring must produce information that allows for this criterion to be tested.

In contrast to target or threshold-based monitoring objectives, long-term (context) monitoring programs must meet the needs of management issues that have yet to be discovered. For example, the BBS, with 50 years of data across the United States and Canada, provides trend information with variable power to detect change for a wide range of bird species, most of which have no stated desired trends or minimum thresholds.

When a monitoring team selects a set of habitat attributes to measure, they must also decide the level of intensity needed for measuring each attribute, because the cost of monitoring will vary, depending on how intensely the attributes are measured. Monitoring intensity is a combination of sample size (sampling intensity), the number of subplots within each sampling unit, the precision with which measurements are made (e.g., ocular estimate, measured to the nearest centimeter), and whether attributes are measured or modeled (section 3.2.6). Some attributes can be sufficiently estimated with existing, remotely sensed data that are relatively inexpensive to track over time. Although other attributes require fairly accurate field measurements, the cost for sampling can be reduced by taking measurements at low sampling intensities over broad spatial extents. For attributes that require estimates with smaller variance and standard errors, the sample size of field plots must be increased. Ultimately, the intensity of the monitoring effort must be sufficient for meeting monitoring objectives and for providing affordable information for management actions.

A monitoring program can reduce costs and obtain sufficiently detailed information by incorporating both low- and high-intensity approaches. Extensive sampling (low spatial intensity) over a broad geographic area can quickly identify the level of variation within the target population and locate areas meriting more intensive sampling. For example, FIA data can provide a comparison of conditions across broad areas to help select areas requiring greater precision of estimates and more spatially intensive sampling. The *Rangeland Ecosystem Analysis and Monitoring Handbook* for the Forest Service Pacific Northwest Region (USDA Forest Service 2006a) provides an example of how sampling methods vary with intensity and scale. Each scale of analysis includes recommended methods for extensive and intensive approaches.

Another example is the use of landtype associations (e.g., midscale; Winthers et al. 2005) to rank areas for more intensive sampling. On the Monongahela National Forest

(West Virginia), managers used low-intensity data at the midscale to rank areas as Indiana myotis (*Myotis sodalis*) habitat. Most of the forest was quickly removed from consideration as bat habitat (DeMeo 2002), which released intensive survey funds for other important survey efforts. Chapter 4 describes approaches for monitoring many common habitat attributes at different levels of intensity and at multiple scales.

Spatial Extent

As a component of the objective, state the geographic area where the monitoring will take place. Not only does this information identify the location, but it also defines the spatial extent so that the monitoring design team can select sampling methods, sampling intensity, and levels of precision that are appropriate for the spatial extent identified. Identifying the spatial extent is also the first step in delineating the sampling frame, the area from which sampling units will be drawn and to which statistical inferences can be made.

Ideally, the sampling frame covers the full spatial extent over which information is needed for the monitoring results to accurately represent the habitat conditions within the area of interest. In reality, however, the sampling frame is frequently a subset of all potential habitats across the defined spatial extent. The full range of potential habitats is the **target population**, and the selected range of habitats to be monitored is the **sampled population.**

For example, managers might select the entire sagebrush (*Artemisia* spp.) community on a national forest as the target population, yet limit the sampled population to units that are of a certain size or proximity (e.g., units in sagebrush stands within 0.5 mile of a road and more than 1.0 acre in size). Statistical inferences are thereby limited to relatively large sagebrush stands close to roads. Another example is the target population, which might be habitat on lands of multiple ownerships that are adjacent to or embedded within National Forest System (NFS) lands. If habitat data on non-NFS lands are available or can be readily obtained, these lands can be included in the sampling frame. If not, the team must recognize that the sampled population does not necessarily represent the target population, and that the study design justifies statistical inferences about habitat quantity or quality only to the sampled population on NFS lands.

Decisions concerning the extent of the sampling frame are particularly important because the sampling frame constrains both current and future sampling. If the monitoring team wanted to change the sampling frame extent at a later time, some of the detected differences in habitat would be the result of sampling a different population. It could be impossible to determine the degree to which the measured change represents a meaningful difference in habitat or is simply an artifact of sampling different populations. Thus, the sampling frame chosen at the beginning of a monitoring program will constrain all future monitoring. For this reason, the careful choice of an appropriate sampling frame is essential to the success of a long-term monitoring effort.

Desired Precision of the Monitoring Information

The desired precision for habitat monitoring (i.e., variance, size of standard errors, and confidence intervals) is established through an iterative process that includes deciding the level of sampling intensity, determining whether existing data will meet the information needs, choosing an appropriate sampling design and method of data analysis, and then using the data analysis method to evaluate the precision of existing or new data with confidence intervals or **power analysis**. Avoid using generic levels (e.g., an estimate that is within 10 percent of the actual value, 90 percent of the time), unless those values are appropriate for the attribute(s) of interest. Nichols and Williams (2006) suggest that monitoring programs be designed to address narrow and well-defined system responses, following the model of hypothesis testing. If this approach is followed, it is much easier to determine meaningful levels of precision of the data. The decision regarding how much precision is needed for each attribute places requirements on the monitoring sampling intensity, because low-intensity sampling methods will, by nature, yield less precision (e.g., larger coefficients of variation, wider confidence intervals, or less power in statistical tests) than more intensive methods.

Because each set of data has its own limits in precision, it is necessary to decide whether monitoring will be accomplished with a specific set of existing data, with new data, or through a combination of the two. When choosing a source of data, recognize that, over time, changes in technology or field methods could greatly affect the original estimate of attainable precision if data collected in the future will no longer be comparable to the original data. If data acquisition might undergo substantial changes between monitoring intervals, consider using another data source, even if the apparent precision at the onset appears to be less.

Sampling designs and analysis methods affect the type of power analysis used for evaluating expected precision and accuracy. For example, trend analyses, repeated measures, and analyses of proportions each use a different type of power analysis (Gibbs and Ene 2010).

Minimum Detectable Change

An effect size or minimum detectable change (Elzinga et al. 1998) should be determined primarily by ecological criteria early in the planning stages of a monitoring study. Knowledge of the expected variability associated with the selected attribute(s) is also helpful for estimating the attainable values of effect size or minimum detectable change, however. Either using pilot data or existing data, examine the width of confidence intervals relative to the desired minimum detectable change on the most important two or three attributes, and use formulas in standard statistics textbooks to evaluate the sample size necessary to produce confidence intervals with acceptable width. If new data are to be collected, estimate the required sample size using procedures described in section 3.3.3 under the subheading, Estimate the Number of Sampling Units Required. One can evaluate the power to detect a given minimum detectable change by a standard statistical test that is appropriate to the analysis method. We recommend evaluating the width of the confidence intervals because that is usually easier to understand than **Type I** and **Type II errors** associated with a power analysis.

For trend detection, use preliminary data and a simple linear regression to see whether a trend is evident as a positive (or negative) slope. Are the sample size and the number of sampling periods sufficient to produce an estimated slope so that the confidence interval on the slope does not contain zero? Equivalently, is adequate power available to detect a certain positive (or negative) slope in the regression line?

If a power analysis or an evaluation of the confidence intervals indicates that the data are unlikely to provide the desired precision or to detect the desired level of change, consider increasing the sample size (section 3.3.3). This augmentation will require collecting new data using the same sampling design and data measurement protocols as the existing data.

Several Web sites provide information to assist with power analyses. To help plan for adequate power to detect a positive (or negative) slope, the Southwest Fisheries Science Center offers TRENDS (Gerrodette 1987), available at http://swfsc.noaa.gov, under the tab Publications/Research software. Another freeware package is MONITOR (Gibbs and Ene 2010), available at http://www.esf.edu/efb/gibbs/monitor/. Lenth (2009) provides software for several forms of power analysis at http://www.stat.uiowa.edu/~rlenth/Power/. Forest Service employees have access to information on power analyses, formulas, and SAS codes on the internal Forest Service statisticians' Web site at http://fsweb.rmrs.fs.fed.us/statistics/statmethods/IntroSASpower.html. (This Web site and other sites beginning with "fsweb" are only available to Forest Service employees or those with access to a Forest Service server.)

After performing sample size calculations, evaluate funding capabilities to determine whether it will be possible to achieve the estimated sample size with available resources. If not, adjust the desired precision to a level that is less optimal, yet capable of providing useful information, to help meet the stated objectives. If it is not possible to obtain precision that is sufficient for the information needs, it may be necessary to abandon plans for using the selected attribute(s) in the monitoring program and to seek alternative attribute(s) whose mean values can be more accurately estimated.

Quantitative Standard for Evaluating Monitoring Results

Ultimately, the purpose of monitoring is to support management objectives and to inform managers when current approaches need to be altered. Therefore, the monitoring objective statement needs to describe the condition or set of conditions expressed as one or more quantitative standards that could trigger a change in management.

For trend monitoring, the standard could be a specified percentage of increase or decrease in the amount or quality of habitat or specific habitat attributes. We recommend selecting a value that is meaningful and observable for the attribute or habitat being monitored. For threshold monitoring, the quantitative standard could be a desired condition statement from the current land management plan or a threshold established in a species' recovery plan. The quantitative standard can also be a historical reference condition (see chapter 6).

3.3.2 Evaluate Whether Existing Information Can Be Used

Two types of existing information are available: field-sampled data and remotely sensed data, such as aerial photography and satellite imagery (see chapter 4 for a thorough review of existing information).

Field-Sampled Data

Field-sampled data are any form of data collected by surveyors on the ground, as opposed to data collected through airborne devices and satellites. The FIA program provides the most comprehensive source of field-sampled vegetation data in the form of a continuous inventory, with broad applicability for monitoring purposes. Other forms of field-sampled data are forest stand exams (including the nationally standardized Common Stand Exam), rangeland inventories (including current vegetation sampling methods), Terrestrial Ecological Unit Inventories, and wildlife habitat surveys.

Evaluate the potential for using field-sampled data using the following steps.

- Determine the spatial extent of existing data and compare it with the spatial extent of the proposed monitoring program.
- Evaluate whether the data were collected with a structured sample design (e.g., simple random, stratified random) or whether potential bias exists from **convenience sampling**.
- Determine whether the existing data include measurements of the desired habitat attributes and, if not, if it will be possible to derive these habitat attributes from other measured variables in the data set.
- Compute confidence intervals (or run a power analysis) on attributes of interest to determine sample size requirements (see subheading, Estimate the Number of Sampling Units Required, in section 3.3.3).
- If the existing data cover a smaller spatial extent than the proposed monitoring program, OR if the existing plot data have less precision than needed for the monitoring objective, design an unbiased probabilistic sampling procedure for increasing the sample size.
- Become familiar with the field protocol used for the existing data so that it can be repeated at existing and new sites. This process includes evaluating whether the field protocol documentation is sufficient to allow for repetition.

Remotely Sensed Data

The Forest Service and other land and resource management agencies use a wide variety of **remotely sensed data**, including traditional aerial photography, moderate- to very-high-resolution digital satellite imagery, regional downscaled climate models, and active systems such as radar and Light Detection and Ranging (**LIDAR)**. Some of these data (discussed in more detail in chapter 4, section 4.5) have been regularly acquired, have been archived for decades, and are available as data layers in a Geographic Information System (GIS).

The process of evaluating whether remotely sensed data can be used for a specific habitat monitoring program is similar to the process of evaluating field-sampled data.

- Determine the spatial extent of existing data and compare it with the spatial extent desired for the monitoring program.
- Determine whether the data source will likely be available over the desired time frame of the monitoring program because rigorous monitoring requires continuity in all aspects of data collection and analysis. If the remote sensing **platform**, **sensor**, or classification algorithms will change across the desired timeframe, using these products for monitoring may not be feasible.
- Determine whether the existing data include the habitat attributes of the emphasis species and, if not, whether it is possible to derive the habitat attributes from other analyzed or sampled variables in the data set.
- Become familiar with the methods and assumptions that went into the classification process.
- Determine whether the **spatial resolution** is appropriate for one or more scales of habitat used by the emphasis species.
- Evaluate map accuracy to ensure it is sufficient for the intended analysis objective (Foody 2002, 2009). If using a subarea of a larger remotely sensed product, try to determine the likely error rates within the subarea for those habitat attributes that are most critical.

3.3.3 Plan and Design for New Data Collection if Needed

Frequently, an inventory or monitoring objective will require acquisition of new field-sampled data when any of the following are true.

- The area of interest is fairly small.
- The habitat attributes cannot be derived from existing data.
- The existing data do not indicate use of an appropriate sample design.
- The sample size of existing data is inadequate to meet monitoring objectives.
- The existing data are useful but not current.

In the last two cases, and assuming those existing data meet minimum standards for bias and precision, the new monitoring program should use the same sampling design and data-collection methods as the existing information to facilitate analyses of the attributes between sampling periods. If the need for new data is not tied to an existing data source, data acquisition will require development of a sampling design (i.e., making decisions about the sampling units and their placements) and development of new standard operating procedures for measuring variables in the field.

Numerous books and articles discuss sampling methods and designs. Introductory texts by authorship include Elzinga et al. (1998), McComb et al. (2010), Morrison et al. (2008), Scheaffer et al. (1990), and Thompson et al. (1998). In addition, Vesely et al. (2006) provide a good introductory presentation of sampling in the context of population monitoring. Sampling methods that account for imperfect detectability include distance sampling (Buckland et al. 2001) and occupancy modeling (Mackenzie et al. 2006). Figure 2.4 of Elzinga et al. (1998) illustrates the process steps for creating a sampling design for quantitative monitoring. The following section follows these authors' steps with special emphasis on details relevant to a habitat monitoring program.

Define the Sampling Unit

Select an appropriate type of **sampling unit** for the attribute(s) that will be measured. If more than one type of sampling unit seems appropriate, use the experience of others to guide in this decision. For example, a fixed-area plot is the typical sampling unit used for measuring plant density, frequency, and biomass (chapter 4, section 4.4.2), whereas a plotless method is the typical sampling unit frequently used to obtain tree basal area. Some attributes are best measured within subsamples of primary sampling units (secondary sampling units or elements). Measurements within secondary sampling units are usually not independent from one another, so the data are generally aggregated into mean values for each primary sampling unit, which serve as the basis for analyses.

Describe Sampling Unit Size and Shape

Before choosing the size, shape, and placement of the sampling units, it is important to become familiar with the spatial distribution of the selected attribute(s) within the area of interest. Most attributes demonstrate some degree of spatial clumping, whether from the topography of the site, species-specific growth patterns, site quality, or site history. The size of the sampling unit should reflect the scale at which the clumped patterns of distribution occur. Nested plot designs are a good approach when monitoring for both common and rare attributes. For example, larger plots might be used for rare elements such as snags, and smaller nested subplots used for more abundant saplings. The larger plots, used for rare elements, help avoid having many sampling units with zero values, which affects data analysis.

As the size of the sampling unit increases, the amount of time required to measure the attributes increases, so a monitoring team should select a sampling unit size that is reasonably small while still reflecting the spatial variability of the attribute. Larger plots have fewer problems with edge effects, however. Ultimately, sampling unit size should be one that yields the highest statistical precision for a given area sampled or a given total amount of available time or money (Elzinga et al. 1998).

The shape and orientation of the sampling units should reflect the spatial distribution of the attribute. The goal is to include as much variability within plots and reduce variability between them to achieve higher precision in the estimated parameter of the habitat attribute.

Determine Method of Sampling Unit Placement

Sampling unit size and shape are influenced by the spatial distribution of measured attributes, but the placement of the sampling units must be made by a probabilistic procedure (e.g., simple random sampling, stratified random sampling, or systematic sampling) and must provide good interspersion throughout the area of interest (Elzinga et al. 1998). Regarding stratification, we reemphasize that "Strata should remain fixed on the landscape over time and data should not be restratified based on some other strata of interest that may arise in the future" (Vesely et al. 2006: 3-11). Stratified random sampling should only stratify on characteristics that will not change during the life of the monitoring program. In particular, one should not stratify on attributes such as current vegetation types and habitat quality that are almost certain to change.

Probabilistic location and good interspersion of units provide the basis for making valid statistical inferences for the study area that will stand the test of time. Problems with lack of independence in systematic sampling plans have been overemphasized in the literature, and methods exist for obtaining relatively unbiased estimates of the sampling errors for systematic sampling (Manly 2001, Stevens and Olsen 2003). Nevertheless, a certain amount of art and professional judgment are involved in designing and analyzing systematic sampling plans. Therefore, we reiterate the importance of involving a statistician in all aspects of monitoring program planning.

Much of this chapter has focused on single-attribute monitoring, but in reality, wildlife habitat is multidimensional and multiscalar, resulting in an interaction of many attributes at several scales. Because of these interactions, it is not possible to select a sampling design and sampling unit placement that are optimal for all attributes. We recommend designing the monitoring program around the two or three attributes judged to be most important to the emphasis species, with the understanding that other attributes will also be measured, although less optimally.

Decide Whether Sampling Units Will Be Permanent or Temporary

Monitoring can be done using either the same sampling units or new ones for each sampling period. The advantage of resampling the same units (repeated measures design) is that it reduces the variance and width of a confidence interval on the slope of a trend line (or other coefficients in more complex models). Similarly, the variance of the difference in means is less when evaluating abrupt changes between sampling periods (section 3.2.4).

Statistical tests for detecting trends or abrupt change between sampling periods are more powerful because the evaluated metric represents the change across time at each sample location rather than change between means of different sampled units.

A disadvantage of repeated measures, however, is the additional expense of marking permanent plots in the first year and the added amount of time required to relocate them during subsequent sampling periods, especially if field personnel change between sampling periods. Moreover, permanent markers are often less than permanent because they can be removed or displaced by humans or damaged by animals and weather. Fortunately, Global Positioning System units have added tremendously to the ease of georeferencing sampling units, but plot corners, centers, and microplots may still need to be physically marked. The use of buried iron, such as pieces of rebar, reduces loss of permanent markers.

An alternative to using permanent plots is to select new sampling units each sampling period, a strategy that may be necessary if visits to a unit tend to bias future measurements owing to trampling or collection of materials. Moreover, if the locations of measured attributes are expected to shift around from year to year (e.g., annual plants), the advantages of permanent plots diminish, and the added cost of marking plots may not be justified.

Selection of new units for each sampling period does not have the inherent bias of a fixed set of units (a fixed set is always greater or less than the true mean for a specific period) and will tend to provide a better estimate of true status of a parameter over time. When the monitoring objective is detection of change or trend in the least amount of time, however, we recommend the repeated measures design.

A blend of these two strategies is the rotating panel design (McDonald 2003). With this approach, the entire set of sampling units is permanent, but only a subsample, known as a panel, is visited every sampling period. All sampling units are eventually sampled over the course of several sampling periods. The rotating panels enable a larger portion of the area of interest to be sampled, whereas the advantages of the repeated measures design come into play after all the sites have been visited for two or more rotations. See McDonald (2003) for a detailed discussion of the issues involved in the analyses of both types of designs. The FIA program uses a rotating panel design with each panel revisited on a 10-year rotation.

Estimate the Number of Sampling Units Required

A statistician with a sense of humor, Doug Johnson, USGS, once said that the answer to the question of how large the sample size should be is that which results in the use of all available funds, an answer that has a lot of truth in it. Certainly, selecting a sample size requires more input than simply using the outcome of a power analysis. Difficulties begin when a sample size that is optimal for one attribute is not necessarily best for other attributes. Also, a sample size based on variation in the past will not necessarily be adequate under a scenario of future variation. For overall sample size estimations, we recommend selecting two or three of the most important attributes and basing the recommended sample size on a compromise for the set. Using these attributes, the recommended approach is to conduct a pilot test and estimate the number of sampling units that will be needed, using the desired width of confidence intervals or the outcome of an appropriate power analysis, as described in Gerrodette (1987) and Morrison et al. (2008).

Next, statistically resample the pilot data with an assumed model for the true status and trend of, for example, the mean response of an attribute over time (Manly 2007). Vary the minimum detectable change according to the model. Resample the **residuals** about the mean in the pilot data, with replacement, to mimic different sample sizes from the population of residuals and observations about the assumed model with a given minimum detectable change. Repeat the process perhaps 1,000 times and determine the number of times the simulated change is detected by an appropriate confidence interval procedure or statistical test. This proportion (say 911/1,000 = 91.1 percent) is the power of a given sample size to detect a given change from an assumed model using the variation evident in the pilot data.

In other words, resample the pilot data as if they are the populations to be observed in the future and determine the likelihood that important minimum detectable change levels can be discerned with different sampling efforts. The advantages of this procedure are that few assumptions must be made, the natural variation observed in real data is mimicked, and computations can be completed on standard desktop computers using add-ons to spreadsheet software. Manly (2007) provides a discussion of resampling and randomization procedures that do not require many of the assumptions in classical parametric statistics.

If a pilot study is not yet available and a rough approximation of sample size is needed, evaluate existing data to estimate the expected variance of the attribute. For example, if the range of measurements on an attribute is expected to be approximately symmetric from 40 to 100 with a mean of about 70, then a good guess for the standard deviation is 10; i.e., an interval about the mean of plus or minus 3 times the standard deviation is expected to contain about 99 percent of the observations. Be aware, however, that this approach involves varying degrees of approximations and assumptions (e.g., a normal distribution) and may yield unsatisfactory estimates of sample size requirements.

Scheaffer et al. (1990) and other statistics texts contain standard formulas for estimating the sample size necessary to achieve a confidence interval with prespecified one-half-width on parameters when comparing data to a threshold value. One of the best recommendations for planning to obtain power to detect a significant trend over time with a simple linear regression line is given by Gerrodette (1987). Also, see the discussion of freeware computer packages TRENDS and MONITOR in section 3.3.1 previously, under the heading, Desired Precision of the Monitoring Information.

Determining the sample size for an inventory or monitoring program is a mixture of art and science. Plan for a compromise of precision on the two or three most important parameters and make adjustments in unit size, unit shape, and methods for each attribute (not all attributes need be measured in association with every sample unit); estimate the variation that will be present in the future, using a pilot study, if possible; estimate the sample sizes to detect important effects; and monitor the monitoring program.

3.3.4 Conduct a Pilot Study

If the monitoring program will be partially or fully based on data collection in the field, we highly recommend conducting a pilot study or series of studies before implementing the full monitoring program. One primary reason for a pilot study is to evaluate variability in the values of each habitat attribute to estimate the sample size needed for achieving the desired precision to detect a given change (as described under Estimate the Number of Sampling Units Required in the previous section). A pilot study, however, can also provide experience and information that will result in a more effective monitoring program.

For example, a pilot study can be used to test out a new protocol or new field equipment, train field personnel, resolve logistical problems related to accessing plots or obtaining measurements, and estimate the amount of time and the cost per sampling unit. Moreover, the outcome of a pilot study can be used to convince funding authorities that the full-scale monitoring program is feasible and practical (van Teijlingen and Hundley 2001).

Depending on the degree to which a particular monitoring protocol is novel, pilot studies will generally involve multiple stages with increasing rigor and formality as the final protocol is approached. For example, at the early stages, limited field testing of protocols can identify a plethora of practical problems. When these problems are systematically eliminated, concerns shift to the statistical reliability of the design. In these latter stages, we recommend that pilot studies be large and formal—dry runs of the actual monitoring program with sufficient sampling intensity to produce initial estimates of variance. Obviously, these later stages are more expensive and time consuming than smaller scale protocol development and will require formal acknowledgment when estimating labor and costs.

3.3.5 Document the Decisions in a Monitoring Plan

The final step in planning and design is to document all decisions and the underlying rationale for these decisions in a monitoring plan. We recommend a modified version of the outline shown in Vesely et al. (2006) for specific inventory and monitoring strategies. Our outline here is essentially the same, with the addition of an Introduction, and with the monitoring objectives embedded in the section on Planning and Design. Topics under each main heading can serve as subheading titles in the plan.

1. **Introduction.** Include the overall goals and objectives for monitoring and explain how they are tied to management goals. Provide rationale for monitoring habitat of an emphasis species or group of species (chapter 2, section 2.3.1).
2. **Planning and Design.** Present a conceptual model for the species or species group, including the levels of habitat selection, habitat requirements, and habitat stressors

(chapter 2, section 2.3.2). Identify the habitat attributes derived from the conceptual model. State the monitoring objectives for these habitat attributes at the appropriate spatial scales. Describe the sampling design, including sources of existing data as well as the spatial and temporal design for collecting new data, either field sampled or remotely sensed.

3. **Data Collection.** State how existing data will be obtained and augmented as necessary. For field-sampled data, describe methods of obtaining measurements of each habitat attribute. For remotely sensed data, describe the process of deriving habitat attributes through image interpretation, classification, or sampling. Include a section on logistics that describes required permits, personnel training, equipment acquisition, contracts for completion of field or GIS work or statistical consulting, and any other relevant logistical information.
4. **Data Storage.** Identify stewards of the monitoring data and state where data will be archived and maintained. (See chapter 9 for a discussion of data storage.)
5. **Data Analysis.** Describe the intended types of data analysis, including map products. Identify the summary statistics that will be used after the initial year, and identify the statistical tests (if any) that could be used to compare monitoring results from two points in time and to evaluate multiyear data.
6. **Reporting.** Identify a reporting schedule (e.g., 1 year, 5 years, 10 years). Include a description of the monitoring reports that will be associated with each reporting interval. State how these reports may inform management decisions.

Chapter 10 of this technical guide provides three examples of habitat monitoring programs. The first two examples systematically demonstrate the process steps that we identify in chapters 2 through 9. The third example is a monitoring plan for mole salamander (Ambystomatidae) habitat, following the outline presented in the preceding paragraphs. The salamander habitat monitoring plan may appear more complex because each main heading is subdivided into three spatial scales of habitat. It is a realistic example, however, that illustrates both the main principles of a monitoring design and the multiscale nature of habitat monitoring.

3.4 Conclusions

A monitoring program is more likely to be successful if time is invested in the initial planning and design. The basic principles of sampling design have been incorporated into numerous references for population monitoring, but this chapter provides the unique role of incorporating these principles into habitat monitoring.

Decisions about habitat monitoring are distinctly different from population monitoring in several ways. For example, population monitoring typically relies on field collection of new population data, whereas habitat monitoring may be able to use data collected from other sources. The ability to use existing data is an advantage, but it brings with it the need for careful consideration regarding whether existing data adequately cover the area of interest, are at an appropriate spatial resolution, and will continue to be collected over the life of the monitoring program. The ability to monitor habitat attributes with remotely sensed data offers advantages as well, but it comes with increased error rates because the attributes are most often modeled, not measured.

Another unique challenge with habitat monitoring is that the team needs to decide whether to monitor attributes individually or to combine them in a model of habitat and then monitor the abundance and quality of that habitat. The decision to combine attributes into a model creates new challenges about model selection and increases uncertainty regarding whether the models adequately represent habitat of the emphasis species. For this reason, we recommend population surveys to evaluate habitat model assumptions (see also chapter 5).

In many ways, however, the basic principles of designing a monitoring program apply equally to population monitoring and habitat monitoring. These principles include stating an objective, selecting indicators (i.e., habitat attributes), deciding on the desired level of precision, deciding on a minimum detectable change that is biologically meaningful, deciding on whether to use independent samples or repeated measures, selecting the appropriate methods of data analysis, and making plans for reporting results and incorporating them into management. Although these principles have been stated elsewhere, they are so essential to any monitoring program that we have reiterated and emphasized them again in this chapter.