Monitoring and Evaluation Plan for Idaho Wildlife Mitigation Projects

Robert S. Unnasch, PhD
Steven K. Rust
David Leptich
Robert C. Martin
Edward Bottum
Gregg Servheen
Mary Terra-Berns

March 2003

Idaho Conservation Data Center
Department of Fish and Game
600 South Walnut, P.O. Box 25
Boise, Idaho 83707

Steven M. Huffaker, Director

Prepared through funding provided by:
Bonneville Power Administration
Project Numbers 199206100 and 19910600

Table of Contents

INTRODUCTION

Background: Congress passed the Northwest Power Planning and Conservation Act (Act) on 5 December 1980. Section 4(h)(10)(A) of the Act directed the Bonneville Power Administration (BPA) "to protect, mitigate, and enhance fish and wildlife to the extent affected by the development and operation of any hydroelectric project of the Columbia River and its tributaries in a manner consistent with the Northwest Power Planning Council's (NWPPC) Fish and Wildlife Program (Program)."

The NWPPC Fish and Wildlife Program addresses the need for monitoring and evaluation (M&E) to ensure that mitigation goals are attained (NWPPC 2000). Section 3.1B (NWPPC 1995) calls for evaluation that "will monitor overall program implementation, evaluate the effectiveness of actions taken, and judge their scientific merits." Section 11.4 (NWPPC 1995) states that the NWPPC is interested in ensuring that mitigation actually occurs on the ground, and accordingly, is providing for monitoring to determine if projected benefits to wildlife result from the Program. The Program calls for an independent scientific review group to evaluate the progress and success of wildlife mitigation efforts (NWPPC 1995, Section 11.4A.2). Consequently, the Independent Scientific Review Panel (ISRP) was formed and made the following recommendation: Monitoring, which was based on habitat units (HUs) determined through habitat evaluation procedure (HEP) analysis, be expanded to include a requirement for some degree of direct monitoring of target (and perhaps some non-target) wildlife populations (III.B.25, ISRP Report 97-1, July 1997). Sponsors of BPA-funded wildlife mitigation projects recognize and support the need for an M&E program that goes beyond HEP (U.S. Fish and Wildlife Service 1980a), is based on good science and standard methodologies, can be applied in an adaptive management context, and balances the need for information with an appropriate level of effort when conducted in a management context. This monitoring and evaluation plan is a response to these Program and project needs.

Monitoring Framework: The scale at which a monitoring program will be applied is a defining consideration in the development of a monitoring program. Spatial scales can be geographic (regional or local), ecological (landscape or habitat), or jurisdictional (Federal, State, Tribal). Biological scales may incorporate entire ecosystems or local populations of a featured species. Temporal scale may consider seasonal, annual, or long-term variability/stability and outputs of a community. An ideal monitoring program would transcend all spatial, biological, and temporal scales. In reality, broad-scale extensive monitoring programs often lack the sensitivity to detect local level perturbations. Conversely, more intensive monitoring methods applicable to research on a site-specific basis rapidly become too costly and labor intensive to apply on a broad scale. This M&E plan attempts to balance both of these needs by identifying portfolios of monitoring sites. These portfolios may include a single large site (e.g., Boundary WMA) or a constellation of smaller sites within a common landscape (e.g., Lake Pend Orielle WMA). In doing so, the plan sacrifices the ability to quantitatively track changes at every site, yet provides a structure that can affordably be implemented over the long-term.

Monitoring can be conducted at three levels of intensity:

1) *Tier I Trend* monitoring is sufficient to answer questions about the trend in population or habitat condition over a broad scale. It has the advantage of being relatively inexpensive to implement. However, its lack of precision makes it relatively insensitive to local conditions or management actions. On a programmatic scale (NWPPC Fish and Wildlife Program) we believe that HEP analysis (U.S. Fish and Wildlife Service 1980a) falls into this category. Particularly for projects that endeavor to mitigate a finite ledger of HUs associated with losses from a specific hydropower project, HEP adequately meets the monitoring needs, at a programmatic level, to ensure mitigation goals are being achieved. Consequently, HEP will remain an integral part of our overall monitoring strategy.

2) *Tier II Quantitative* monitoring is able to answer questions about population trends, community diversity, and species relative abundance in the context of local habitat condition or management action. Although more costly to implement, this level of monitoring has sufficient sensitivity, and defined levels of confidence, to provide feedback on management actions in an adaptive management context. Additionally, by collecting site-specific data according to standardized protocols these data may be used across multiple spatial and biological scales. Consequently, they may contribute data points to regional, national, or international monitoring efforts. Conversely, by collecting data that contributes and are comparable to a broader data set the manager can better interpret results (e.g., declines in amphibian populations as a local versus more general biological problem). Most of the methods outlined in this M&E plan fall into this level of monitoring. A purposeful effort was made to select methods that are widely employed in field biology or to adopt appropriate monitoring protocols from national monitoring programs to maximize the utility of the data collected. A significant limit of this level of monitoring intensity is that it is not designed to evaluate the causes of change in habitat or population trends.

3) *Tier III Research* monitoring is the most sensitive level of monitoring. At this level we are able to answer questions about causal relationships between specific habitat attributes and population demographic parameters. The data demands to achieve the statistical power to answer these types of questions make this the most expensive level of monitoring to employ on a per area basis. This is beyond the management context of this M&E plan. However, if Tier II Quantitative monitoring suggests a management problem that cannot be adequately addressed by a review of the literature and through the manager's experience, nothing in this M&E plan constrains a manager from developing a site-specific monitoring program at this intensity level to address specific problems.

Both *Tier II* and *Tier III* monitoring requires a rigorous sampling to allow for unbiased estimates of change. Insufficient or inappropriate sampling design most often results in data whose value is questionable and which are unable to detect biologically significant changes.

Estimating temporal trends in plant and animal populations is a common goal of biologists and managers. While, on rare occasions an entire population can be counted or measured, trends are typically inferred from counts of individuals made on samples over time. Trends represent the sustained patterns in count data that occur independently of cycles, seasonal variations, and irregular fluctuations in counts.

A common problem in trend detection is that sources of "noise" in counts obscure the "signal" associated with ongoing trends. The probability that a monitoring program will detect a trend in sample counts when the trend is occurring, despite the "noise" in the count data, represents its *statistical power*. Although statistical power is central to every monitoring effort, it is rarely assessed. The consequences of ignoring it include wasting resources by collecting data in excess of what is needed to meet the goals of the program, or conversely, collecting data which are insufficient to make reliable inferences about population trends. A more complete discussion of statistical inference and statistical power is included in Appendix A.

<u>Project and Monitoring Goals</u>: Monitoring and evaluation consists of assessing changes in habitats, populations, or communities that test the effectiveness of management measures. Adaptive management is the process of using scientific information to evaluate and improve management decisions. Conceptually, adaptive management is based on the need to maintain operational flexibility to respond to monitoring and research findings. Hence, adaptive management is the practical application that links monitoring and management. The goal of any monitoring program is to provide information that verifies whether management objectives are being met. Therefore, monitoring goals are dependent on management goals. For each BPA-funded wildlife mitigation project, site-specific management objectives will be identified and incorporated in the project's M&E plan.

In support of each project's management goals, the objectives of this monitoring and evaluation plan include the following:

1. Track progress toward full mitigation of the HU losses identified in hydroelectric project impact assessments.

2. Evaluate the success or failure of management activities by:

   a. Periodic mapping of the vegetation to assess changes in the distribution and extent of all plant communities at each site.

   b. Periodic reassessment of the number of habitat units at each mitigation site. (HEP monitoring empowers the interpretation of changes in targeted wildlife populations.)

   c. Monitoring trends in the composition and structure of targeted plant communities.

   d. Monitoring population parameters (relative abundance, distribution, and population trends) of selected target and non-target land birds and waterfowl species.

   e. Maintain a photographic record for permanent monitoring sample points.

3. Ensure that all monitoring data collected are sufficient to detect a 2.5 percent annual change over the span of ten years with a statistical power of 80 percent. This annual rate of change equates to a total change of 20 percent from starting conditions after the ten-year period.

4. Adopt standardized monitoring methodologies that are compatible with monitoring at larger scales and the scientific literature. This will maximize the usefulness of the data collected within the NWPPC Fish and Wildlife Program as well as at regional or national scales.

This monitoring protocol has been designed to assess changes in inter-related components of the ecosystem. Vegetation mapping provides a foundation upon which the HEP monitoring can be interpreted. Vegetation monitoring allows for an assessment of changes within the mapped vegetation cover types or wildlife habitats. HEP monitoring allows for extensive, qualitative assessments of habitat quality. Finally, avian monitoring relates changes in cover, HUs, and plant community composition and structure directly to the management targets: wildlife.

Sampling protocols for bald eagles, small mammals, and herptofauna were considered for in the early development of this monitoring plan but were ultimately dropped. Bald eagle nest and roost sites are currently monitored through other well developed and successful programs (e.g., Beals and Melquist 2001). Small mammal and herptofauna monitoring procedures were dropped from the plan as the effort required to acquire suitable levels of statistical confidence in population trends was greater than the potential benefit of achieving mitigation program monitoring goals and objectives.

Protocols for monitoring noxious weed populations are describe in the following text. These activities, however, will be conducted outside the scope of the monitoring program. For this reason, monitoring frequencies and costs have not been fully developed.

PROGRAM SAMPLING DESIGN

Introduction: This wildlife and habitat monitoring program is designed to provide managers with information on population and community trends through time that can be used in an adaptive management context. Monitoring is an ongoing obligation of management and should itself be viewed as an adaptive process. The BPA HU ledger has not been fully mitigated. Most of the land base that will eventually be managed and monitored is not currently identified. Without good knowledge of the total land base, distribution, juxtaposition, block size, and condition (degree of restoration required) of mitigation properties, it is difficult to design an efficient monitoring program that anticipates all future needs. However, the current design is explicitly flexible and allows for (1) the addition of new monitoring stations as lands are added to the portfolio and, (2) the reduction of the sampling frequency at individual monitoring stations to relieve the pressure on resources of adding new lands and additional stations.

The long-term monitoring database for this project will be developed through both observational and quantitative monitoring. Observational monitoring includes the use of such things as photo plots and incidental wildlife observations that may suggest changes in plant or wildlife communities at a qualitative level. These data have the advantage of being relatively inexpensive to obtain but are limited because they depend on subjective interpretation. Quantitative monitoring depends on actual measurement of population or community attributes, and these data are amenable to statistical analysis. The primary disadvantage of quantitative monitoring is that it can be expensive and time consuming. However, quantitative monitoring can provide estimates of direction and magnitude of change before change is grossly evident, is less biased than observational monitoring, and is the most objective way to evaluate the success of our mitigation and management programs. Appendix A provides a primer on quantitative assessment and statistical inference.

Areas of Statistical Inference: The monitoring program will be applied at a spatial scale that attempts to balance broad-scale, extensive and site-specific, intensive monitoring needs by identifying portfolios, or groups, of monitoring sites. These groups may encompass a single large site (e.g., Boundary) or a constellation of smaller sites within a common landscape (e.g., the sites surrounding Lake Pend Orielle). These groups of sites constitute the area over which statistical inference will be made. In defining areas of statistical inference in this manner, the plan sacrifices the ability to quantitatively track changes at every site, yet provides a structure that can affordably be implemented over the long-term. General areas of inference for current mitigation lands are shown in Figure 1. With the addition of lands within the Wildlife Mitigation Program, the exact delineation of these areas of statistical inference will be re-evaluated and possibly expanded to accommodate changes.

VEGETATION MAPPING PROCEDURES

Vegetation mapping will occur at a minimum map scale of 1:24,000 and a preferred scale of 1:12,000. A 30 meter minimum mapping resolution will be maintained for geographical information system raster data files. A minimum mapping unit of two acres within this 30 meter digital resolution will be maintained. Mapping projects may draw on color aerial photography, digital orthoquads, or Landsat imagery depending on the spatial scale of the site under investigation. Mapping projects will accurately display the following base vegetation attributes: potential natural vegetation, existing vegetative cover, vegetation structure, and ecological condition. Vegetation classification schemes for each of these four attributes will comply with those maintained by Idaho Conservation Data Center (2003). A minimum of 25 percent of the polygons identified will be ground truthed using standard plant community ecology methods.

HABITAT EVALUATION PROCEDURES

Introduction: The Habitat Evaluation Procedure was developed in 1980 by the U.S. Fish and Wildlife Service (USFWS 1980a; USFWS 1980b; USFWS 1981). HEP uses a species-habitat based approach to impact assessment, and is a convenient tool to document the predicted effects of proposed management actions. The NWPPC endorsed the use of HEP in its Columbia River Basin Fish and Wildlife Program to evaluate wildlife benefits and impacts associated with the development and operation of the Federal Columbia River Basin hydroelectric system. The HEP was used to evaluate wildlife habitat impacts attributable to habitat inundation at the following federal hydroelectric projects in Idaho: Palisades, Minidoka, Anderson Ranch, Black Canyon, Deadwood, Boise Diversion, Albeni Falls, and Dworshak.

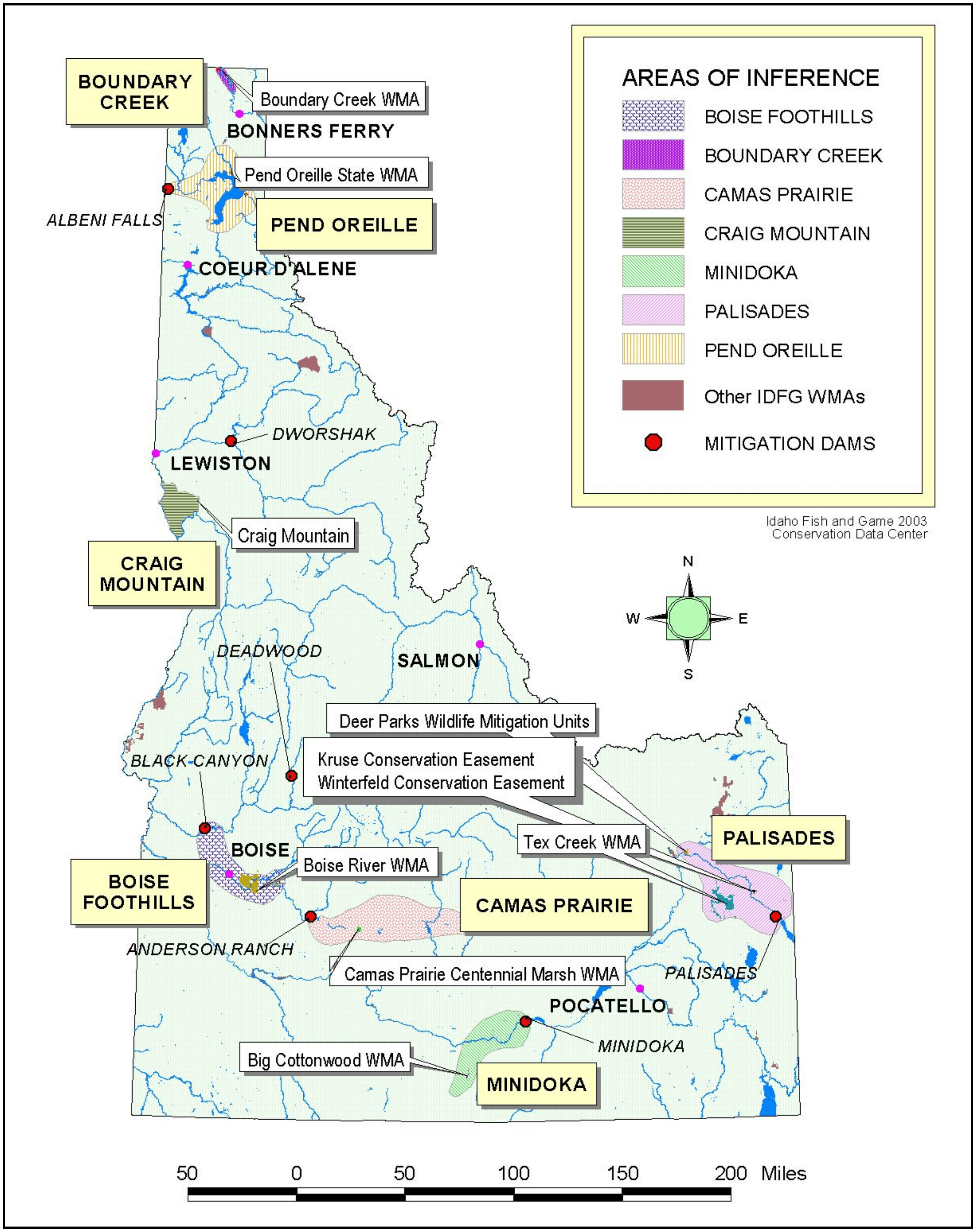

Figure 1. Areas of Inference. White boxes point to existing Wildlife Management Areas. Multiple Wildlife Management Areas combined on the basis of related mitigation efforts and regional proximity are Areas of Inference, illustrated in light yellow.

The objective of using HEP in an M&E program is two-fold. First, it provides an objective and quantitative assessment of wildlife mitigation credits property acquisition projects. Secondly, the baseline HEP evaluation describes existing ecological conditions (limiting factors) on properties and may be used to assist managers in developing future management activities.
Methods: The HEP is based on the assumption that habitat for a selected species can be described by a Habitat Suitability Index (HSI). This value is derived by evaluating the ability of key habitat components (e.g., hiding cover, snag density, forage availability) to supply the life requisites of selected wildlife species. Habitat quality, expressed as the Habitat Suitability Index, measures how suitable the habitat is for a particular species when compared to optimum habitat. The HSI varies from 0.0 to1.0 (optimal). The value of an area to a given wildlife target species is the product of the size of that area and the quality of the area for the species. This product is comparable to "habitat value" and is expressed as a habitat unit (HU). For a particular target species, one HU is equivalent to one acre of optimal habitat (HSI=1.0). Target species are used in HEP to quantify habitat suitability and determine changes in the number of HUs supported by a particular area. Consequently, a HEP assessment is only directly applicable to the target species selected. The degree to which predicted effects can be extrapolated to a larger segment of the wildlife community depends on careful species selection (USFWS 1980b). Target species and their HSI models selected for HEP analyses in the M&E program would generally be those target species and models used during hydroelectric project wildlife impact assessments. Likewise, field- and remote-sampling methods would generally follow those used during the wildlife impact assessments. During field sampling, transects are lengthened or occasionally shortened to achieve a 90 percent confidence level for our parameter point estimates. Adequacy of habitat sampling is determined using the formula (Zar 1984):

$$\frac{z^2 \text{ x } s^2}{e^2}$$

Where:

z= the critical normal value (p=0.1) from any standard statistical reference

s= standard deviation

e= tolerable error level

Data Analysis: Habitat cover types are outlined on aerial photographs and a planimeter or dot grid is used to estimate the total acreage of each cover type. Geographical information systems (GIS) will be used to estimate total acreage of each cover type when accurate data layers are available. The habitat units for each target species in each cover type are calculated using the formula:

HU= (cover type area) X (HSI value).

Published and modified HSI models are used in this analysis. Where published models are modified to better reflect local conditions, modifications must meet U.S. Fish and Wildlife Service standards (USFWS 1981). Habitat units are tabulated across target species and cover types to get total HUs for each species and each cover type for the property.

The NWPPC Fish and Wildlife Program requires that a baseline HEP analysis be completed within two years of acquisition of a mitigation property and every five years thereafter. This schedule will be followed as part of the ongoing M&E efforts on this project. Some acquisitions are primarily to protect existing high-quality habitats, where management is largely custodial and significant increases in HUs are not anticipated. Other acquisitions require extensive restoration, and substantial gains in HUs are the expected outcome. Results of HEP analysis must be interpreted in this context. For the purposes of adaptive management, we expect to maintain, within the limits of normal temporal variability, at least the baseline number of HUs on every property. A 20 percent drop in baseline HUs would trigger a management response.

QUANTITATIVE MONITORING PROTOCOLS

Using the Universal Transverse Mercator (UTM) coordinate system, a permanent grid with spacing of 200 meters or less will be established on each mitigation property. Grid points will be sequentially numbered and represent potential monitoring sample points that can be randomly selected by use of a random numbers generator. The 200-m spacing is equal to the preferred sample point separation for land bird point-count stations (Huff et al. 2000), and yields one potential sample point for every 4 ha of habitat. Closer grid-point spacing decreases the probability that data from adjacent sample points are independent and increases the risk of double counting birds when using variable-radius point-count sampling techniques in particular.

Drawing the sample of points to be monitored is complicated by the fact that we are still in the implementation phase, and additional properties will be added on an annual basis for the next 10+ years. The sampling scheme must be cost effective, provide a data set that provides a long-term perspective on meeting management objectives, and is flexible enough to incorporate new properties as they are acquired. Consideration must also be given to the fact that habitats do not occur in equal proportions and that some habitats are intact while others require restoration. Taking these concerns into consideration we have devised the following sampling scheme:

Sampling will be done with an intensity appropriate to meet the monitoring goals identified above (80 percent power to detect a 2.5 percent annual rate of change). As additional properties are purchased, additional permanent sample points will be identified to ensure that these goals are maintained. Thirty (30) sample points will be established within each site constellation (or area of statistical inference, Figure 1). Sample points will be re-sampled on a three-year rotation. This is a sufficient frequency to capture long-term trends in plant and animal population and community change (See Appendix B).

A random sample of long-term monitoring sample points will be drawn from all possible sample points. Once identified as part of the sample to be monitored, these points will become part of a permanent subset of points to be used for long-term monitoring.

This random sampling design makes no "*a priori*" distinction between sample points that fall on intact habitat where management is custodial and restoration sites where the management is active, and community changes may be dramatic even in a short amount of time. At a programmatic and project scale this is appropriate to document the success or failure of conservation strategies from a long-term monitoring perspective. However, it may not provide managers with adequate feedback on the success of site-specific management prescriptions. Managers may choose to supplement this basic sampling scheme with additional sample points randomly selected from within a site-specific prescription area for assessing trends at a finer spatial scale or for Tier III Research. These supplemental sample points will not become part of the long-term permanent sample-point set. They may be revisited more frequently than every three years and/or dropped from monitoring altogether at any time at the manager's discretion. The sample size recommendations resulting from the power analyses reported here (Appendix A) hold true for multiple spatial scales. Thus, for example, if a manager wishes to monitor changes in Townsend's warbler at a single site, s/he will need a minimum of 30 census points. If s/he is interested in the same question across a constellation of sites across a landscape, 30 points randomly located throughout that landscape will suffice. In the latter scenario, however, s/he will not be able to draw any statistical inference for the individual sites.

Monitoring in an adaptive management context implies benchmarks or desired outcomes against which management success can be measured. The vegetative and wildlife community structure of intact terrestrial habitats can act as a benchmark for the effectiveness of restoration management. A subset of permanent sample points will be retrospectively (that is after the random sample has been drawn) identified from each habitat to serve as reference sites against which restoration management may be evaluated. Additional reference sites, both within and outside of the project boundaries, may be subjectively identified to secure a minimum of three reference sites for each habitat. Sample points selected as reference sites will initially be sampled for three consecutive years to establish a strong baseline data set. Based on initial results, permanent baseline monitoring plots may also be established (to the extent possible) within formally designated ecological reference areas (e.g., USDA Forest Service Research Natural Areas) that are located in areas adjacent to mitigation properties but are functionally independent of mitigation properties and associated management. When available and applicable the scientific literature will provide an additional source of reference benchmarks for project evaluation. The data analyses proposed in this report work both for assessing change over time and for comparison between managed areas and reference sites (Provencher et al. 2002).

LAND BIRDS

Introduction: Birds are important components of biological diversity in most ecosystems. Monitoring the health and long-term stability of bird communities can provide an important measure of overall environmental health (Morrison 1986). Birds are good environmental monitors for several reasons: many species can be monitored simultaneously with a single method, methods for monitoring are well understood and standardized, birds occupy all habitat types, and as a community represent several trophic levels and habitat use guilds. Monitoring species abundance, community diversity, and trends provides information that can be used to determine the effectiveness of management actions in moving towards conservation goals.

Perhaps more than any other species or community proposed for monitoring, land birds present the opportunity for standardized data collection that can be incorporated into national monitoring programs. Dovetailing our monitoring efforts with national monitoring efforts can be important in interpreting the results of our monitoring efforts. Many species of birds are neo-tropical migrants whose populations are effected by factors remote from the data collection point. Standardized methods allow for recognition of declines in abundance or diversity as a local phenomenon (triggering a change in local management) or a broader scale phenomenon that does not necessarily implicate failed management at the local level. The goals for this monitoring program are, first to be able to detect a 2.5 percent annual rate of change in the abundance of land birds, and second to be able to assess changes in the community composition and structure.

Sampling Methods: Point counts will be used to monitor land birds on this project. Point counts are the most widely used quantitative method used for monitoring land birds and involve an observer recording birds from a single point for a standardized time period (Ralph et al. 1995). The methods follow the recommendations of Ralph et al. (1995) and are consistent with the methods employed by the USDA Forest Service, Northern Region land bird monitoring project (Hutto et al. 2001) and recommendations for the Idaho Partners in Flight Bird Monitoring Plan (Leukering et al. 2000). A preliminary power analysis of land bird monitoring data is presented in Appendix B.

A minimum of thirty randomly selected permanent census points must be established in order to detect the desired change in population numbers (see Appendix B for a preliminary power analysis). A ten-minute point count will be conducted at each point. All points will be visited a minimum of three times during the breeding season (mid-May to early July) with a minimum of seven days between counts. Point counts should be started at 15 minutes after official sunrise and completed by 10:00 a.m. Weather conditions should be warm and calm enough for bird detection by sight or sound.

Fixed-radius plots (where the radius is arbitrarily small) reduce the interspecific difference in delectability by assuming that: (1) all the birds within the fixed radius are detectable; (2) observers do not actively attract or repel birds; and (3) birds do not move into or out of the fix-radius during the counting period. This allows for comparisons of abundance among species. Unlimited radius plots maximize the amount of data collected because they include all detections and are appropriate when the objective is to monitor population changes within a single population (Ralph et al. 1995). If population estimation is desired, then additional distance data must be collected. However, density estimation is beyond the scope of this monitoring plan. Additional information on establishing point count stations, data collection, and sample data forms can be found by referencing Ralph et al. (1993; 1995) and Huff et al. (2000).

Protocol for Data Collection:

Point setup. Birds should be tallied in two distance bands: one 0-50 meters from the point center and one >50 meters from the point center. It is critical that every detectable individual within 50 meters of the observer be recorded, as this will maximize statistical power and allow for interspecies analyses. Sample points should be identified with a numbered pin flag, whose location has been recorded using a GPS. This enables rapid relocation of points during subsequent visits and provides sufficient data to position these points in a GIS system. It is critical that the 50-meter threshold is consistently applied, and thus it is recommended that five to six markers, 50 meters distant from the sample station, be placed at in clear view of the observers. These markers significantly reduce observer bias by ensuring that all observers share a "true" 50-meter threshold.

Data Acquisition. Upon reaching a sample station the observer should stand quietly for five minutes prior to recording bird activity. During that time, s/he should record information on the time of day, air temperature, wind speed, sky conditions, and noise index (accepted indices and values are shown on the sample data sheet, Appendix C).

Observers should record the species, sex, and age of all individual birds seen or heard during the 10-minute count period following the initial 5-minute quiescent period. Species observed flying, but not alighting within the observed area, should not be counted. During the count, data should be recorded in three time periods (0-3 minutes, 3-5 minutes, and 5-10 minutes). This will allow the data to be partitioned or pooled for comparison to the U.S. Fish and Wildlife breeding bird survey data, research data reported in the literature that commonly use 5-minute point counts, and 10-minute point count data recommended and collected by national bird monitoring programs. Field observers should be highly qualified to detect birds by sight and sound. A field data sheet is provided for avian monitoring in Appendix C.

Data Analysis: Data can be pooled in a variety of ways; however, it is important to realize that the minimum sample size necessary to confidently detect the desired 2.5 percent annual rate of change is 30 points. Smaller samples will be unable to statistically separate long-term trends for sampling "noise" (see Appendix A). The mean number of detections per point (by species) within a cover type will be used as a *qualitative* measure of species abundance which may empower the interpretation of larger trends. Abundance across cover types, within a land management unit, will be expressed as 90 percent confidence intervals of the individual cover-type data pooled across the land management unit and weighted by the proportionate areal extent of each cover type. Trend analysis on species population abundance data will be done by simple regression analysis (Zar 1984). Any regression analysis will not be conducted with less than 5 data points. However, it is inadvisable to examine each species separately, as this inflates the probability of committing a "false-change" error; that is, seeing changes where none, in reality, exist. Multi-species analyses should be done using Analysis of Covariance.

A major goal of the land bird monitoring program is to detect changes within the community. Given the analytical problems in examining individual species changes in aggregate it is best to evaluate the community as a whole. Changes in the avian community will be assessed using the squared Euclidean distance method of Ratliff and Mori (1993) and/or the Bray-Curtis dissimilarity index (Krebs 1989). Both methods examine all species in aggregate, and hence community structure and composition. Unlike the Shannon-Wiener and Jaccard's coefficients, neither Bray-Curtis nor the squared Euclidean distance are effected by zero values for some species.

Species richness will also be used as a measure of diversity. The species list will be developed and supplemented with incidental sightings from throughout the year.

WATERFOWL

Introduction: Waterfowl are comprised of a diverse group of birds with widely different habitat needs for survival and recruitment. Some goose populations have expanded in the face of extensive national wetland losses. Conversely many duck species, which are less terrestrial and more dependent on wetland quality and availability, have experienced substantial population declines. The Canada goose, mallard, and redhead duck are BPA target species that were used in the HEP analysis habitat loss assessment. Waterfowl breeding-pair and brood surveys are conducted to provide trend data for local breeding populations. Our survey protocols are modeled after waterfowl production survey methods developed and used by the U.S. Fish and Wildlife Service (Hammond 1970, Dan Pennington, Kootenai National Wildlife Refuge, pers. comm.).

Methods: Open water areas and associated wetlands within monitoring areas of inference (Figure 1) will be surveyed annually. Four different types of waterfowl production surveys will be conducted: goose breeding pair counts, goose brood counts, duck breeding pair counts, and duck brood counts. Because of differences in nesting phenology between geese and ducks some different surveys may be conducted concurrently on the same visit to a site (e.g. goose brood counts concurrent with duck pair counts). Surveys will be conducted as a combination of observation point counts, walk/wade surveys, and boat and motor runs as appropriate for the landscape.

Observation point counts are used where there is good visibility, especially from elevated positions, to observe open water areas. When using observation points, disturbance must be kept to a minimum. Observation points are best conducted with the aid of a spotting scope. After data are gathered via observation points a walk/wade survey may need to be conducted to observe additional open water areas that are not visible from observation points.

Walk/wade surveys are best applied to wetlands with shorelines having little emergent vegetation and can be walked efficiently. Small wetlands should be approached carefully and quietly because the broods of some species (especially mallards and pintails) may move overland to avoid detection by the observer. When properly conducted a high proportion of all broods may be seen with this method.

Boat and motor runs are most efficient on open shorelines. Two observers will see more birds than one observer will. However, a single observer is generally a more efficient use of manpower. Consequently, a single observer will always be used to minimize variability in the trend data. Boat speed should be moderate (5-10 mph) and consistent throughout the survey, stopping only to count broods or identify species.

Survey Timing and Frequency: Counts will be completed within the three-hour periods beginning either 15 minutes after sunrise or ending 15 minutes before sunset. Wade/walk surveys may be conducted throughout the day. All surveys will be conducted as close as practicable to the identified target dates for data consistency. Surveys should be conducted when temperatures are moderate and wind speeds are less than 10 mph. Excessive wind moves birds into protected areas. If practical, rain should be avoided.

Goose breeding pair surveys are conducted twice, once each on or near April 15th and May 2nd. Goose brood counts are conducted twice, once each on or near May 16th and June 6th. Goose brood surveys will be done in conjunction with second duck breeding-pair survey and the first duck brood survey.

Duck breeding-pair surveys will be conducted twice, once on or near May 2 for early nesters, and once on or near May 16 for late nesters. Although some protocols call for only two duck brood sampling periods. Three sampling periods provide a more adequate index than two sampling periods. Three duck brood surveys will be conducted on or near June 6, June 28, and July 26.

For waterfowl pair-counts the species and number of pairs should be recorded. For ducks both paired ducks and lone males representing indicated pairs should be tabulated for all species. During brood counts the observer should record species, number in brood, and the age class of the brood. Data will be summarized by species and land management unit and reported annually. Long-term local trends will be monitored against the national waterfowl surveys.

VEGETATION

Introduction: Vegetation provides habitat for most fish and wildlife species. The primary issues regarding the conservation and restoration of vegetation and wildlife habitats are plant community composition, structure, and ecosystem function. Through a number of studies the targeted vegetation has been classified on the basis of composition and structure into plant associations and community types. Plant associations and community types provide groupings of similarity in composition and structure. Several different plant associations or community types may be present within each of these broad cover types. While the vegetation mapping protocol presented above will allow for assessing coarse changes in cover types, the following methods are appropriate for monitoring plant community composition, structure, and ecosystem function within these three broad cover types.

Sampling Methods: Vegetation sampling occurs within square, 4 hectare (200 x 200 m.) cells identified as the area surrounding and centered on each random sample point. The sample protocol consists of three components: (1) coarse-scale composition and structure, (2) fine-scale tree stem data, (3) fixed area, time-constrained species inventory, and (4) fine-scale species composition. Coarse-scale composition and structure is sampled on two 200 m. transects by measuring the boundary between each plant association (using classifications developed by Cooper et al. 1991; Jankovsky-Jones 1997; or more recent), covertype, and structural class (modified from Hall et al. 1995) (see Appendix D). One transect each are placed perpendicular to one another running through the sample point center (Rust et al. 2003).

Tree composition and stem density data are collected on nested 0.04 ha. (11.3 m radius) and 0.1 ha (17.8 m radius) circular plots. Live tree stems, standing dead stems, and logs (of sapling size or greater, see Appendix D) are tallied by species, size class, and (where appropriate) decay class. Large live tree stems (> 20.9 inches diameter at 4.5 feet), snags, and logs are tallied by species on the 17.8 m. radius circular plot.

A 20-minute time-constrained plant species inventory is conducted on the entire 0.1 hectare plot all species that are not otherwise recorded (in the stem density data or on the point transect) are listed. Combining species encountered during sampling and this time-constrained search should result in a comprehensive inventory of vascular (and to the extent possible, non-vascular) plant species present within each sampling plot.

Changes in the plant species composition will be assessed by measuring relative changes in cover. A 0.6-hectare circular plot is placed at each of the nine intersections of the 200 x 200 m. coarse-scale sampling grid. Within the 0.6-hectare circular plot six 22-meter transects originating at the center point are established at 60 degree intervals (Figure 2). It is critical that the tape is pulled tight, and is as close to the ground as possible (i.e., thread the tape under shrubs as necessary). Data collection begins at the2-meter mark along each transect. This procedure (i.e., excluding the center 12.5 m^2 of the plot) alleviates concerns for spatial auto-correlation typical with circular plots (J. Herrick personal communication; R. Unnasch, personal observation). At 50 cm. intervals along each transect a pin is slowly lowered from a height of 1.5 meters. Every species touched by the point is recorded. Forty points are recorded along each transect - 240 points within each plot. Species cover is then calculated as:

$$Hi/T*100$$

Where H = the number of 'hits' of species I and T = the total number of pins dropped (in this case T=240).

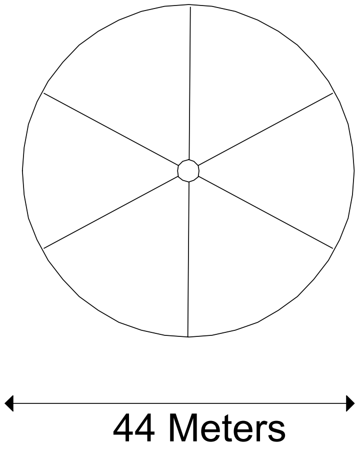

Figure 2. Vegetation sampling plot. Shrub and herbaceous cover is measured along each of the six transects originating from the center of the plot.

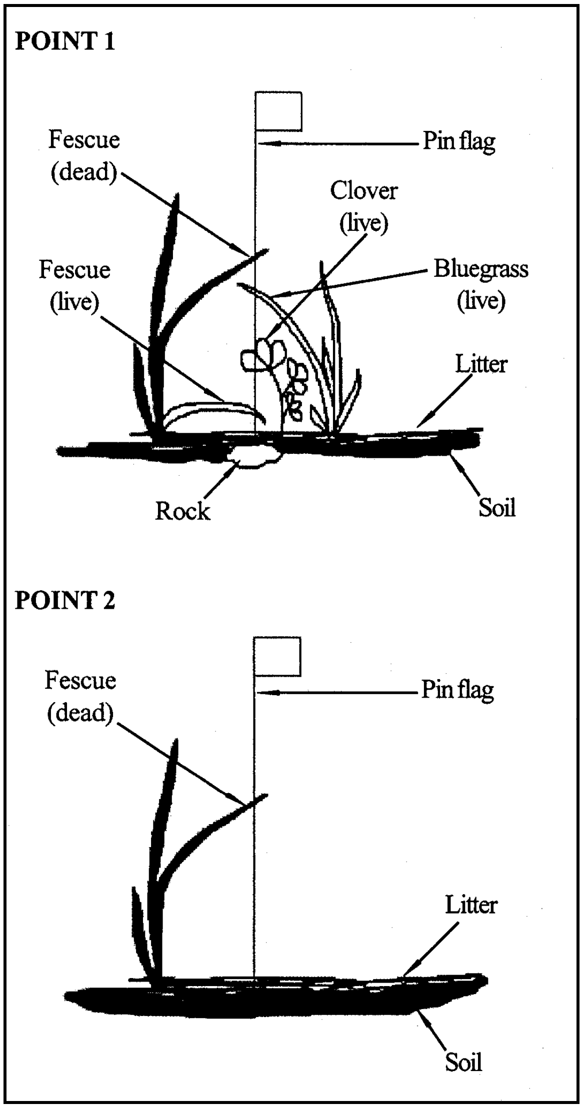

Figure 3. Data recording for line-intercept data (from Herrick et al. 2002).

Data from this example (Figure 3) are entered as follows:

Point	Top Canopy Species	Lower Canopy Species					Soil Surface
1	bluegrass	clover	fescue	L			R
2	~~fescue~~	L					L
3							
4							
5							
etc.							

Reading Points: The observer records the species code for every species intercepted by the pin. Each species is recorded only once per point, even if that species is intercepted several times. If the pin only intercepts dead parts of a plant, a line is placed over the species code. If both live and dead parts of a species are intercepted, only the live species is recorded. The appropriate code for the substrate surface is recorded. When vegetative litter is encountered in a layer above the substrate surface, it is recorded independent of the substrate. Finally, observers should record the code for the substrate surface. An example of how to read line-point data is shown in Figure 3.

Substrate Surface Codes

Code	Description
EL	Embedded litter (if there is no clear boundary between litter and soil
L	Litter
LC	Lichen crust on soil (lichen on rock is recorded "R")
M	Moss
R	Rock (>5mm)
S	Soil that is unprotect by any of the above
W	Woody debris (>5mm in diameter)

A data sheet for recording vegetation data is included in Appendix D.

Data Analysis: This method allows for a sensitive **within plot** tracking of abundance of those species with a cover of 10 percent or greater. Appendix E contains 90 percent binomial confidence intervals for all frequencies when 240 points are sampled. Changes in the abundance of individual species can be assessed either graphically, plotting the confidence intervals over time, or with the use of contingency analyses.

When >7 plots are sampled, it also allows for the use of the Bray-Curtis dissimilarity index, or squared Euclidean distance, to assess changes within the community as a whole (Krebs 1989; Ratliff and Mori 1993).

NOXIOUS WEEDS

Introduction: Noxious weeds are aggressive plants that are not native to an area. They frequently create a large monoculture of themselves. Noxious weeds degrade wildlife habitat; can choke streams and waterways; crowd out native beneficial plants; create fire hazards; poison humans, wildlife, or livestock; and foul recreational sites for use. The spread of noxious weeds can signal the decline of entire ecological watersheds (Morishita and Lass 1999). Noxious weed law requires landowners to control noxious weeds on their land. Control of noxious weeds is consistent with the management objective of all BPA mitigation projects to restore and maintain native habitats.

Methods: Effectiveness of noxious weed management will be tracked by providing estimates of total area of noxious weed invasion and percent cover of noxious weeds by species. Ocular estimation will be used to determine cover by species in five cover class categories: 0 - 20, 21 - 40, 41 - 60, 61 - 80, and 81 - 100 percent. A 1.0 by 0.5 meter sampling frame may be used to aid in cover estimation. Within large patches multiple frames (≥10) should be read and the mean and 95 percent confidence interval width should be reported. GPS mapping will be used to calculate the area of large (>1 hectare) areas of weed invasion. Alternatively, if these areas are sprayed and the spray equipment has the ability to calculate total area treated, this will be an acceptable area estimate. Smaller (1-hectare) areas of weed invasion may be mapped with GPS or by ocular estimation. A noxious weed monitoring database developed for use with a GPS equipped Palm Pilot has been developed by The Nature Conservancy and the Bureau of Land Management. This software links directly to Environmental Systems Research Institute's (ESRI) ArcView software and allows for the rapid aggregation and analysis of this noxious weed data. This easily used database is readily available.

PHOTO POINTS

Although qualitative, photographic documentation of habitat change as it occurs over time can provide an intuitive and compelling record of that change. This record can be especially effective for relating a project's effect to administrators or the public who more easily identify with a picture than a theoretical mathematical function of community diversity. Consequently, a photographic record will be established for each long-term monitoring sample point. One or more photographs will be taken in the direction of each of the four cardinal compass directions at each permanent sample point during its triennial monitoring visit. Photographs will be cataloged and archived for future reference. A digital camera will be used for documenting photo points to simplify archiving and reproduction for reports and presentations.

MONITORING FREQUENCY

A major misconception is that monitoring must be done annually in order to be effective. In reality, annual monitoring is almost always not needed and doing so can result in a significant waste of resources. The following are recommended sampling intervals for the four categories of data presented above:

1. Vegetation cover mapping should be repeated every five years at each mitigation property. The mapping should be done so that the maps are available for the HEP assessments.

2. HEP assessments should be done every five years as required by the NWPPC.

3. Vegetation sampling should be completed every five years at each sample point.

4. Avian sampling should initially be repeated every three years, but these intervals can be extended to every five years after the first 12 years. It is recommended here that avian sampling should be done in years 0, 3, 6, 9, 12, 16, 20, 25, 30, 35, 40, and every five years thereafter.

6. Waterfowl sampling will occur annually.

5. Photo points will be retaken in each year that avian or vegetation sampling occurs.

REPORTING

<u>Habitat Evaluation Procedures</u>: The NWPPC Fish and Wildlife Program requires that HEP analysis be conducted on each acquisition at five-year intervals. This has been the backbone of the NWPPC monitoring and evaluation program to date. No change in reporting procedures for HEP analysis obligations is proposed. Each project leader member will submit HEP reports for properties under their ownership/management at the required time interval under a separate cover as a stand-alone document.

<u>Expanded Monitoring and Evaluation</u>: Permanent long-term monitoring sample sites are visited as presented above. A monitoring and evaluation report that describes the current year's monitoring activities and summarizes finding will be submitted annually. A complete analysis of these data, including

population trend and community analyses, will be included in each periodic report.

We have intentionally designed some flexibility into the program to make it adaptable to the needs and constraints of the local manager. Consequently, it will be important for the core data sets coming from each site be in a compatible format so that these data can be easily and appropriately combined for overall project evaluation and reporting. Use of the sampling methods and the data sheets provided here for the core data sets will ensure this compatibility.

Supplemental Reporting: Where appropriate, efforts to augment this monitoring and evaluation plan to address site specific problems or management actions may occur. Supplemental reports will be written as stand-alone documents and attached to the annual report as an appendix.

COSTS

Land Bird Sampling: Each sampling station takes 25 minutes (+/-) including travel time between nearby sites. Thus, a single person could sample 10 stations in a morning. If there is intersite travel involved (i.e., among sites within a constellation), that number will be reduced to six sites per day. This estimate is based on the Albeni Falls sites, which are an average of 20 km. apart. Thus, a conservative estimate would be four days (mornings) to sample the necessary 30 stations, times three visits per year, or 12 person days per year to sample one large site, or a constellation of smaller sites. It is important to remember that these sites are revisited every three years initially, and every five years later on. Thus, one person working one month each year would cover six sites/constellations every three year cycle.

Waterfowl: Point sampling methods will require at least three hours. Each sample point is visited six times a year. Rounding for travel and application of other sampling methods, this is approximately three 8-hour days per sample point. Approximately 45 points will be sampled annually.

Vegetation Sampling: Vegetation sampling is best done by teams of two - one observer and a recorder. It is estimated that each sample plot would take approximately 3.5 hours to complete. Thus, each team would be able to complete two plots per day. Coarse scale transects will require approximately 3.5 hours to complete with a team of two.

HEP Sampling: Costs of conducting HEP sampling will be evaluated on a site by site basis.

Vegetation Mapping: Mapping procedures will require two to five days for mitigation property less than 240 acres. An additional two days is required for each additional increment of 240 acres. Mapping procedures occur independent of vegetation monitoring, though some monitoring data may assist with mapping efforts.

Photopoints: Photopoints should take an average of one day per site.

Noxious Weed Mapping: This is a part of standard management activities and will not be charged to a monitoring budget.

Estimated Total: A summary of the cost of total cost of the M&E program follows. The estimate of days per year is based on the costs and sampling frequencies identified above and the assumption that, if the plan is applied to current mitigation properties, a total of 210 sample points will be visited during a 3 - 5 year period.

Sampling Element	Field Sampling			Reporting	Total
	days/point	points/year	days/year	days/year	days/year
Land birds	0.4	70	28	30	58
Waterfowl	3.0	45	135	30	165
Vegetation - detailed plot	0.9	42	38	30	68
Vegetation - coarse transect	0.9	42	38	30	68
Vegetation mapping	current approximately 133 days total on five year cycle		30	30	60
Photo point			4	2	6
Total days per year:					425

LITERATURE CITED

Beals, J., and W. Melquist. 2001. Idaho bald eagle nesting report 2000. Idaho Department of Fish and Game, Boise. 16 pp.

Cooper, S. V., K. E. Neiman, and D. W. Roberts. 1991. Forest habitat types of northern Idaho: a second approximation. USDA Forest Service General Technical Report INT-236. Intermountain Research Station, Ogden. 143 pp.

Gibbs, J.P., S. Droege, and P. Eagle. 1998. Monitoring populations of plants and animals. BioScience 48:935-940.

Hall, F. C., L. Bryant, R. Clausnitzer, K. Geier-Hayes, R. Keane, J. Kertis, A. Shlisky, and R. Steele. 1995. Definitions and codes for seral status and structure of vegetation. General Technical Report PNW-GTR-363. USDA Forest Service, Pacific Northwest Research Station, Portland OR. 39 pp.

Hammond, M. C. 1970. Waterfowl brood survey manual. U. S. Fish and Wildlife Service, Washington D.C. 43 pp.

Hayes, J. P., and R. J. Steidl. 1997. Statistical power analysis and amphibian population trends. Conservation Biology 11(1): 273-275

Herrick, J.E., J.W. Van Zee, K.M. Havstad, and W.G. Whitford. 2002. Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems. USDA Jornada Experimental Range.

Huff, M. H., K. A. Bettinger, H. L. Ferguson, M. J. Brown, and B. Altman. 2000. A habitat-based point-count protocol for terrestrial birds, emphasizing Washington and Oregon. Gen. Tech. Rep. PNW-GTR-501. Portland, OR: U.S. Department of Agriculture, Forest service, Pacific Northwest Research Station. 39 pp.

Hutto, R. L., J. Hoffland, and J. S. Young. 2001. USDA Forest Service Northern region landbird monitoring project field methods 2001 west side monitoring. University of Montana Division of Biological Sciences, Missoula, MT. 25 pp.

Idaho Conservation Data Center. 2003. Plant community element tracking database. Electronic database file in Biological and Conservation Data System. Idaho Department of Fish and Game, Conservation Data Center, Boise.

Jankovsky-Jones, M. 1997. Conservation strategy for Northern Idaho wetlands. Conservation Data Center, Idaho Department of Fish and Game. 35 pp. plus appendices.

Krebs, C. 1989. Ecological Methodolgy. Harper Collins, New York.

Leukering, T., D. Faulkner, and M. Carter. 2000. Monitoring Idaho's birds: a plan for count-based monitoring. Colorado Bird Observatory, Brighton, CO. 23 pp.

Morishita, D. W. and L. W. Lass. 1999. Idaho's noxious weeds. University of Idaho Cooperative Extension. Moscow, ID 74 pp.

Morrison, M. L. 1986. Bird populations as indicators of environmental change. Pp. 429-451 in R. J. Johnston (ed.), Current Ornithology, Vol 3. Plenum Press, New York. 522 pp.

Northwest Power Planning Council. 1995. Columbia River fish and wildlife program, 1995 amendments. NWPPC, Portland, OR.

Northwest Power Planning Council. 2000. Columbia River fish and wildlife program. NWPPC, Portland, OR.

O'Connell, T.J., L. E. Jackson, and R. P. Brooks. 1998. The bird community index: a tool for assessing biotic integrity in the mid-atlantic highlands. Report No. 98-4, Penn State Cooperative Wetlands Center, Forest Resources Laboratory, Pennsylvania State University, University Park, PA 16802.

Provencher, L. et al. 2002. Longleaf pine restoration project final report. Report to the US Air Force, Eglin Air Force Base, Florida. The Nature Conservancy.

Ralph, C. J., G. R. Geupel, P. Pyle, T. E. Martin, and D. F. DeSante. 1993. Handbook of field methods for monitoring landbirds. Gen. Tech. Rep. PSW-GTR-144, Albany, CA: Pacific Southwest Research Station, Forest service, U.S. Department of Agriculture; 41 pp.

Ralph, C. J., J. R. Sauer, and S. Droege. 1995. Monitoring bird populations by point counts. Gen. Tech. Rep. PSW-GTR-149, Albany, CA: Pacific Southwest Research Station, Forest service, U.S. Department of Agriculture; 187 pp.

Ratliff, R.D. and S.R. Mori. 1993. Squared Euclidean distancd: A statistical test to evaluate plant community change. USDA Forest Service REs. Note PSW-RN-416.

Rust, S. K., J. J. Miller, and C. Coulter. 2003. Vegetation monitoring on Idaho Department of Fish and Game fish and wildlife mitigation lands: Pilots studies on the Albeni Falls Wildlife Mitigation Project, Pend Orielle and Boundary Creek Wildlife Management Areas. Idaho Department of Fish and Game, Conservation Data Center, Boise. 33 pp.

Sokal, R. R. and F. J. Rohlf. 1982. Biometry. 2nd Edition. W. H. Freeman and Company, San Francisco. 859 pp.

U.S. Fish and Wildlife Service. 1980a. Habitat evaluation procedures (HEP). Ecological Services Manual 102. Division of Ecological Services, Washington D.C.

U.S. Fish and Wildlife Service. 1980b. Habitat as a basis for environmental assessment. Ecological Services Manual 101. Division of Ecological Services, Washington D.C.

U.S. Fish and Wildlife Service. 1981. Standards for the development of habitat suitability index models for use in HEP. Ecological Services Manual 103. U.S. Fish and Wildlife Service Division of Ecological Services, Washington D.C.

Zar, J. H. 1984. Biostatistical Analysis. Prentice Hall, Englewood Cliffs, NJ. 718 pp.

APPENDIX A: Basics of Statistical Inference

Population of Inference: The ultimate goal of any monitoring program is to detect real changes over time within a defined statistical population. In certain cases the entire population can be counted or measured. In these rare circumstances any changes measured reflect real changes in the population. The bald eagle monitoring program herein is an example of such a case. More commonly, however, the population cannot be measured in it entirety, and inferences about population changes must be based upon measurements made on only a small sample. These inferences are subject to errors purely as a result of this need to sample. These sampling errors can be generally classified into two categories - spatial and spatio-temporal. Any effective monitoring program must accommodate for these errors. These accommodations include appropriate sample size, appropriate sample unit size and shape, and appropriate sampling intensity.

Spatial errors derive from the fact that biological populations and ecosystems vary over space. When samples are based upon observations based upon ephemeral plots (i.e., quadrats, transects, or observation points are not permanently located), there is some probability that differences seen among years reflects spatial and not temporal differences.

Spatio-temporal errors result from segments of the populations changing at different rates, or even in different directions. Breeding anurans, for example often increase in numbers in some ponds while simultaneously decreasing in other, nearby ponds.

A common problem in trend detection is that these sources of "noise" obscure the "signal" associated with ongoing trends. The probability that a monitoring program will detect a trend in sample counts when the trend is occurring, despite the "noise" in the count data, represents its *statistical power*. Although statistical power is central to every monitoring effort, it is rarely assessed. Consequences of ignoring it include collection of count data insufficient to make reliable inferences about population trends, or collection of data in excess of what is needed to accomplish the program goals. Monitoring which is ignorant of these issues is fated to result in data that are un-interpretable and conclusions that are suspect. Too often the results of a long-term monitoring program are reported in the vein of “the statistical analysis of our data don’t show the changes we observed.” In reality, this is a tacit admission that the monitoring program was poorly designed initially, and that significant resources were wasted over several years collecting meaningless data.

An explicit goal of this monitoring plan is to prevent such errors from occurring by defining an acceptable statistical power “*a priori*,” and assuring that all protocols are sufficient to meet this criteria. Throughout this preliminary analyses we set our monitoring goals as an 80 percent likelihood (power = 80 percent) of detecting a real trend of 2.5 percent per year over the span of 10 years. This annual rate of change equates to a total of a 20 percent change over ten years. The chance of incorrectly concluding that this change occurred was set at 0.1. Thus, in statistical terms alpha was set at 0.1, and beta at 0.2.

Setting alpha at 0.1 runs somewhat contrary to the researcher’s norm of alpha = 0.05. The reason this higher alpha or ‘false change’ error is the relationship between alpha and beta, the ‘missed change’ error. Choosing to set alpha at 0.5 significantly reduces beta (Figure 3). Because the goal of monitoring is to detect real changes, we want to maximize our ability to detect changes, even if this increases the probability of concluding that change has occurred even when it hasn’t.

In statistical parlance, power is defined as 1 – β (beta), with β being the probability of wrongly accepting the null hypothesis when it is actually false, known as a Type II error. Power, the complement of this event, is the desirable likelihood of correctly rejecting the null hypothesis. In the simplest terms, the power is the probability that a quantitative monitoring program will detect a given change, if that change occurs. Ideally, one wishes to maximize statistical power while minimizing the resources expended.

In a monitoring program, this outcome is influenced by many factors, including count variability over space and time, survey length, number of study plots, within-year effort, chosen magnitude of trend to be

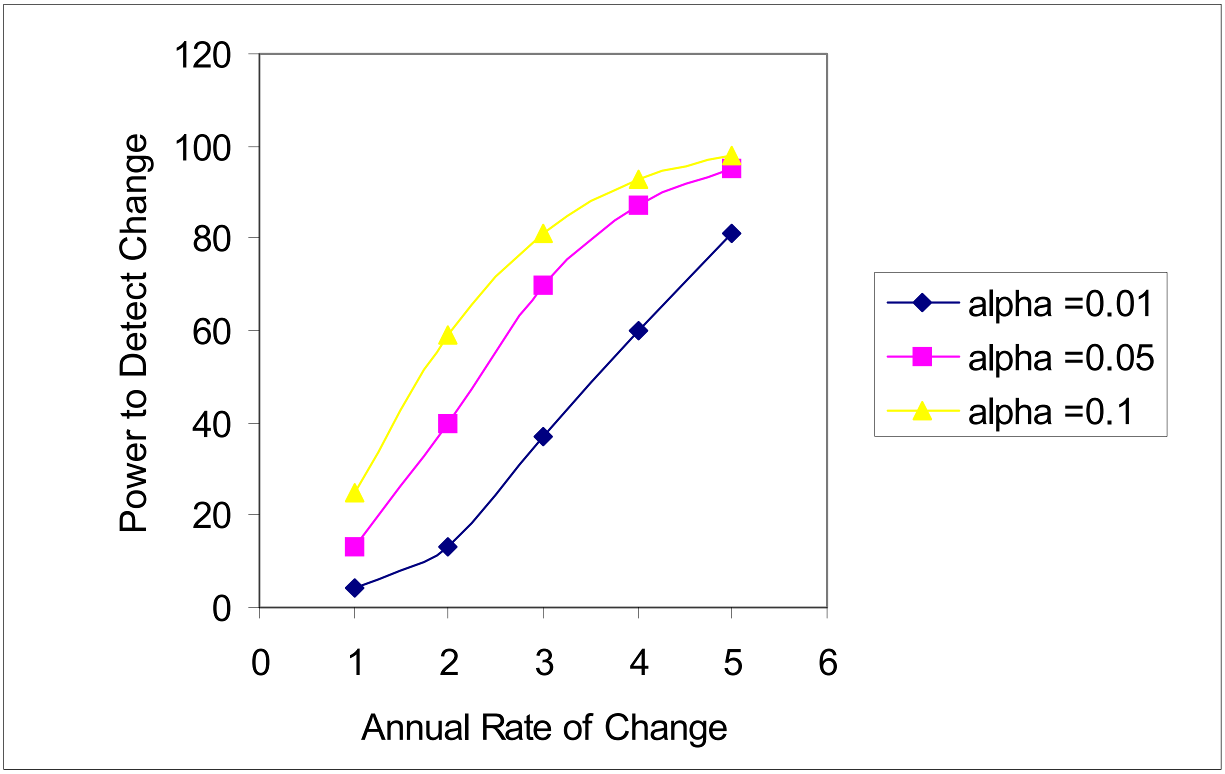

Table 4. The relationship between the power to detect change and the rate of false change error (α).

detected (effect size), and statistical level of significance (alpha (α)—the probability of mistakenly rejecting the null hypothesis, known as a Type I error). To a lesser or greater extent, many of these factors are predetermined, reflecting program goals and defining the monitoring program itself.

Sensitivity of trend detection **(effect size)** is a good example of a variable often driven by management objectives. What constitutes a "biologically significant" rate of population change? There continues to be a large debate on this question, but suffice it to say that the values seen in the literature tend to range between 2.5-5 percent (declines in particular) per year, which translate into 20–40 percent over 10 years. It is clear the stronger the trend, the less effort it takes to detect. Conversely, smaller changes require greater effort to detect. Most monitoring programs have to balance the detectable change with resources available.

The power players. The sampling effort required to achieve a given level of power are affected by sampling design variables. Different programs will have different areas of flexibility where they can boost power. Size of refuge, management plan, and scope of question are just a few variables that will lock in the values of some variables. What follows is a (simplified) discussion of these variables within the context of biological monitoring programs.

Variables that must be defined *a priori* by program designer:

Type I error Level (α). The hegemony has been intact since Fisher decided that a 1 in 20 probability characterizes a rare event for a specific experiment. However, Fisher also noted that Type I error rate must be balanced with other sources of error. While assigning a less stringent alpha runs contrary to tradition, it allows for greater power at no additional cost in resources. Doing so increases the risk of "crying wolf." Throughout this plan we set α = 0.1.

Type II error Level (β). A common criterion for power is 80 percent (β = 0.20) (e.g., Hayes & Steidl 1997), and this was chosen as appropriate for the analyses in this plan.

Time (number of years). The time span to detect change was set at 10 years and longer.

"Tailedness." All analyses were performed assuming declines. This was most conservative, as growth is easier to detect and hence would require smaller sample sizes.

Effect size (trend magnitude). Throughout this report, we set the desired rate of change at 2.5 percent per year for 10 years. This equates to a 20 percent change overall.

<u>Variables that are determined from preliminary sampling</u>:

Count variance over space. When available this was taken from preliminary sample data. When preliminary data were not available, simulated data were drawn from a normal distribution using parameters from the published literature.

Count variance over short periods of time. When available this was taken from preliminary sample data. When preliminary data were not available, simulated data were drawn from a normal distribution using parameters from the published literature.

<u>Variables that are calculated based upon the above considerations</u>:

Number of counts per sampling period.

Number of plots.

In this plan we estimated the statistical power of population the various monitoring programs relative to (1) the number of points, or plots monitored, (2) the seasonal variation of counts per plot, (3) count variation over space, (4) the duration of monitoring, (5) the interval of monitoring, (6) the magnitude and nature of ongoing population trends, and several other factors. Because these factors interact in complex ways to determine the capacity of a monitoring program to detect trends in populations, such basic questions of "how many plots should I monitor" or "how often should I conduct surveys" rarely have intuitive answers.

To address these questions, we utilized the software program MONITOR which uses Monte Carlo procedures to generate many, simulated sets of count data based on a monitoring program defined by the user and sample counts drawn at random from distributions defined by the user. Through multiple trials, MONITOR evaluates how often a monitoring program detects trends of varying strength that actually occur in the count data.

APPENDIX B: Land Birds.

Introduction: When monitoring long-term trends in animal populations and communities, one needs to account for both spatial and short-term (intra-seasonal) temporal variation. Unlike plants that "sit there and wait to be counted," animals are more detectable on some days than others. Thus, since the goal of this monitoring program is to detect real changes over the span of years to decades, we need to account for this short-term variation.

No preliminary count data were available for land birds and so a power analysis was performed based upon simulated observations drawn from information obtained from the published literature. Gibbs et al. (1998) reported that a review of 73 studies revealed an average annual coefficient of variation of 0.6. Based on that information, a preliminary "sample" was created by drawing observations from a normally distributed population with a mean of 10 and an annual, per census point, coefficient of variation of 0.6. A power analysis was performed on these data to provide some guidance on sample size and sampling frequency.

POWER ANALYSIS

There are three considerations when designing a monitoring program for animal populations and animal communities: the numbers of observations per season, the number of samples, and the time between sampling. These will be dealt with in order.

Number of Observations per Season: Because counts of animal populations vary with time of day, weather, and season, it is important to account for this variation by sampling multiple times per year. As illustrated by

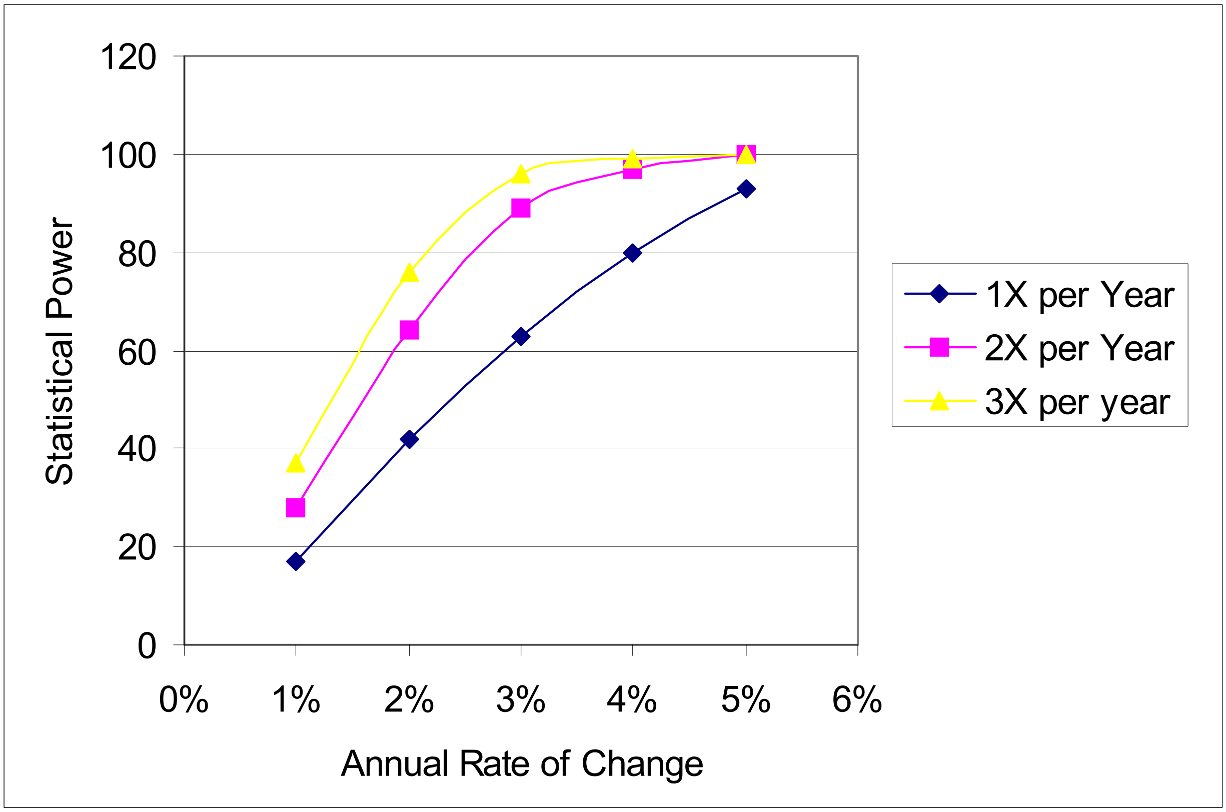

Figure 5. The relationship between statistical power and repeated annual sampling. These results are based upon 3,000 monte carlo simulations using sample data with a mean of 10 and an intra-annual coefficient of variation of 0.6. Each census point was visited every year for 10 years.

Figure 4, multiple observations significantly increase the power to detect long-term trends. This figure represents an analysis examining the increase in statistical power by sampling 25 plots once, twice, three times each year for 10 years. In a statistical sense, multiple samples per year partitions the within-year variation from the longer term trend, increasing the power to detect long-term change. Sampling once per season provides less than a 50:50 chance of detecting a 2 percent annual decline over 10 years; worse than if one simply flipped a coin – heads it changed, tails it didn't. Sampling twice a season increases the power to 64 percent, and three times per year increases it again to 76 percent.

Number of Samples: Many monitoring programs fail to achieve their goals because of insufficient sample size. The realization that sampling intensity was inappropriate often occurs after several years of data have been collected, and the researcher perceives a change but the data do not support that perception. This preliminary power analysis is explicitly designed to prevent that scenario. Here, an analysis was performed on the simulated data set to provide some guidance as to the minimum sample size sufficient to meet the goals of the program (2.5 percent annual change over 10 years, $\alpha = 0.1$; $\beta = 0.2$). One thousand monte carlo simulations each of sample sizes of 10, 15, 20, 25, and 30 were run, with each census point sampled three times per season. The results of this analysis are presented in Figure 5. These results indicate that to achieve a power of 80 percent or better, one needs to collect data from at least 20 census points.

Frequency of Sampling: The above analyses assume that the population would be sampled annually. However, there is potentially a large cost savings if sampling can be done at intervals greater than every year. The following figure examines the relationship between statistical power and sampling frequency. It is clear from this analysis that there is relatively little additional gain in statistical power by sampling annually (Figure 6). The statistical power to detect a 3 percent annual rate of change sampling every year is 90 percent over

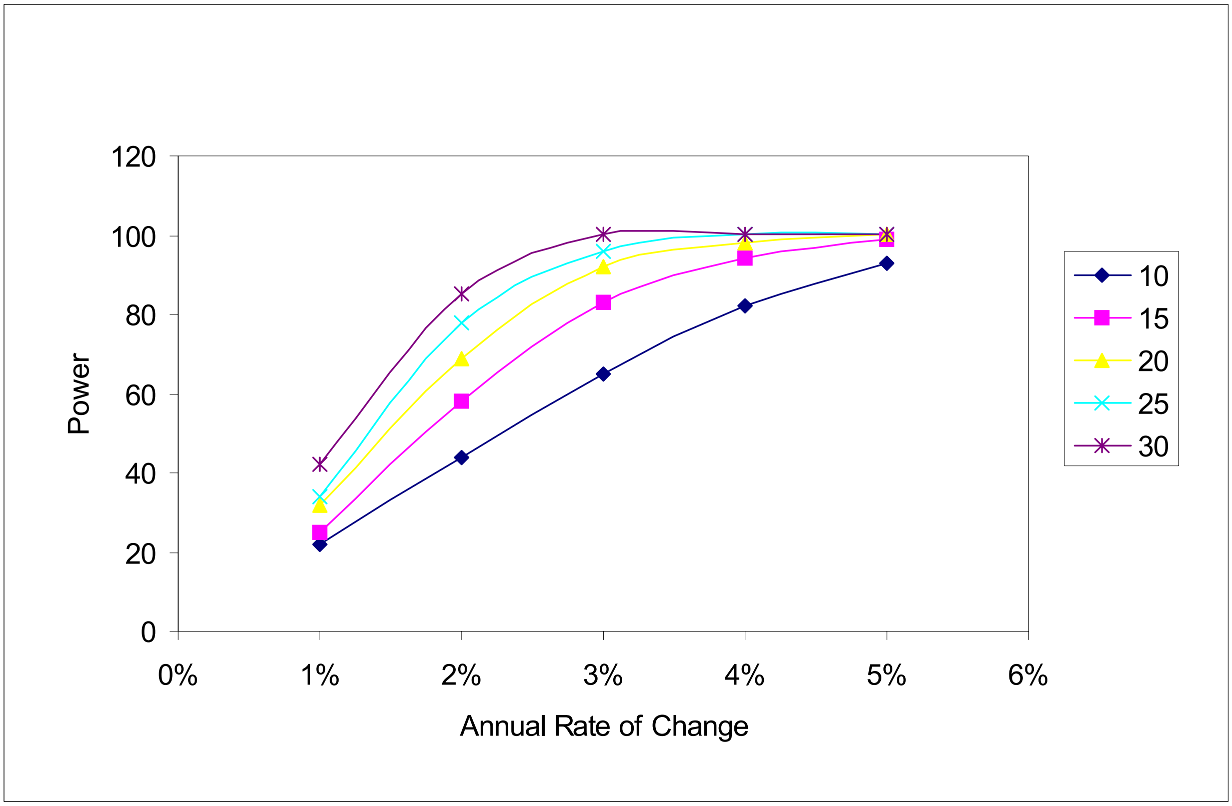

Figure 6. The relationship between statistical power and sample size. These results are based upon 5,000 monte carlo simulations using sample data with a mean of 10 and an intra-annual coefficient of variation of 0.6. Each census point was counted three times per year, every year. Each simulation was set to run for a 10-year span.

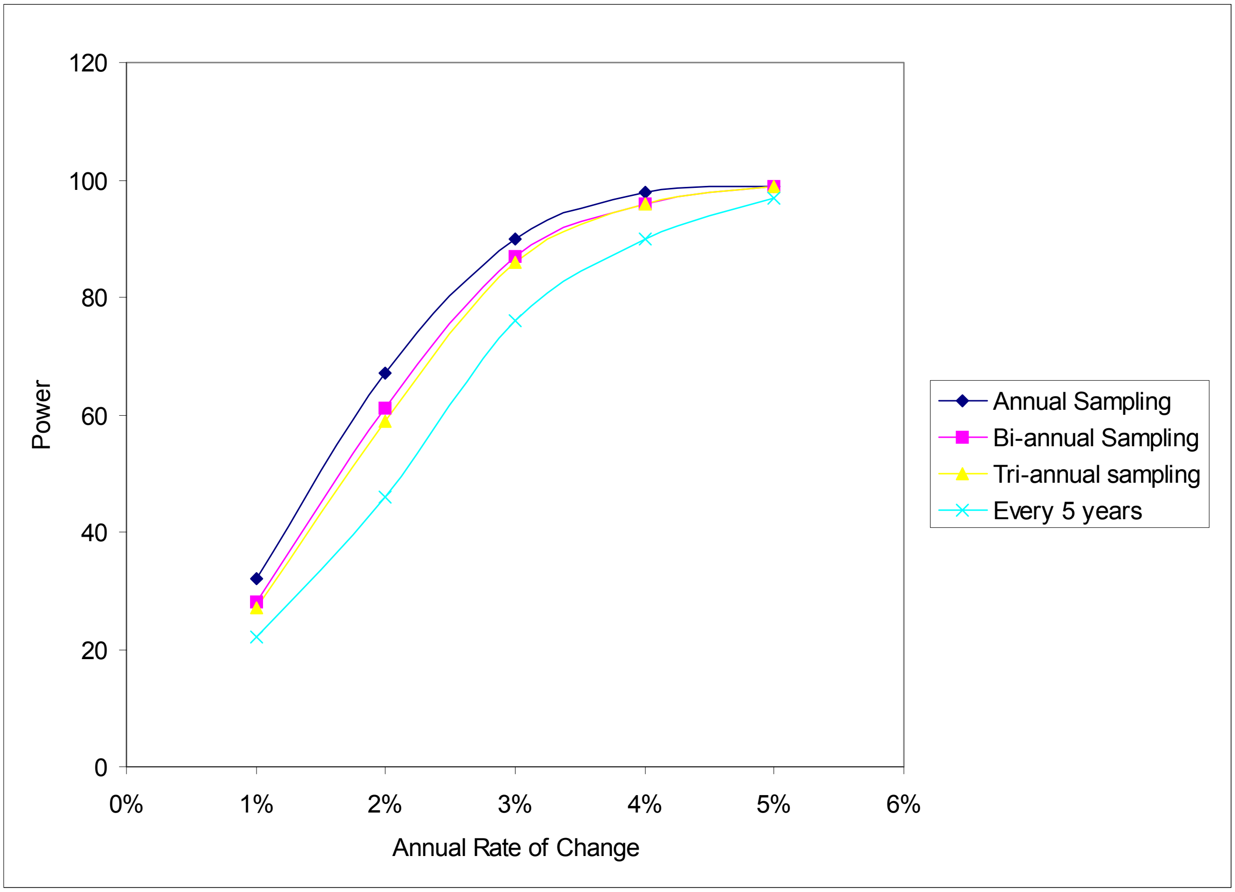

Figure 7. Relationship between statistical power and sampling frequency. These results are based upon 4,000 monte carlo simulations using sample data with a mean of 10 and an intra-annual coefficient of variation of 0.6. Each census point was counted three times per year, every year. Each simulation was set to run for a 10-year span.

10 years, whereas sampling every 2 years results in a minor loss of power to 87 percent, and every three years a power of 86 percent. There is a significant loss of power when sampling is done every fifth year.

Conclusions and Recommendations: **Based upon the above analyses, it is clear that the optimal sampling intensity is to sample 30 census points, 3 times per year, at 3-year intervals.** This sampling regime would achieve the goals of the program, >80 percent power to detect a 2.5 percent annual change over ten years, with the minimum number of points sampled. Specifically, the power of this regime is 83 percent. Although adding additional census points will increase statistical power initially, this additional benefit will decline over time.

Through the power analysis (above) twenty census points was identified as the appropriate sample size. The power analysis assumed sampling would occur three times a year, every year. Sampling every year, however, is not the most efficient use of time, as this sampling frequency does not provide substantial improvement in analytical power (Figure 6). We face a trade-off between 20 samples every year versus a few additional samples (30) each sampling period, but less frequent sampling (e.g., every three years). The later alternative results in a lower sampling effort over all. For example, over a four year period, data are collected at 80 census points (20 points/year) in the former alternative. In the later alternative data are collected at 60 census points (30 in year 1 and 30 in year 4). Thus, the later alternative results in a 25 percent savings in sampling effort while achieving equivalent power to detect significant change.

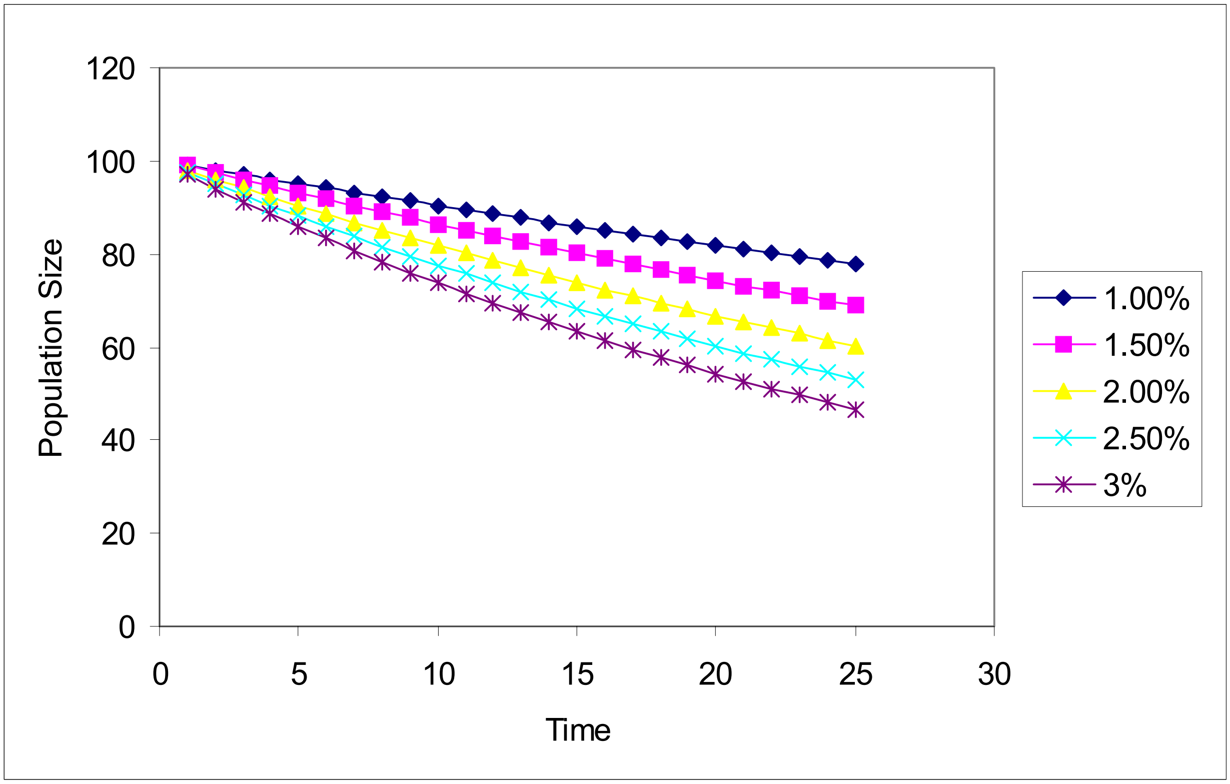

Figure 8. Population changes under various rates of decline.

As data are accumulated, it will be possible to reduce the future sampling intensity. This is primarily the result of two factors. First, the absolute magnitude of change becomes greater (Figure 7), and secondly the within year and spatial variation become better defined. For example, after the fifth sample (year 12) it will be possible to reduce the sampling intensity to 2 times per year, or to reduce the number of points sampled from 30 to 20 and still maintain the required 80 percent power level.

<u>Data Analyses</u>: Assessing changes in the abundance of individual species should be done using simple regression analysis. Care must be taken when performing multiple univariate tests, as alpha becomes inflated as the number of tests increases (Sokal and Rohlf 1982), and if a test of overall change is desired then all species should be tested simultaneously using an ANCOVA.

There are several ways that changes in community structure can be measured. These include the Bird Community Index proposed by O'Connell et al. (1998), similarity measures, and multivariate analyses.

The Shannon-Wiener information index is often proposed as a standard measure of community change. It is, however, a poor choice for many reasons. The most important of these is that it confounds changes in species richness with changes in evenness (Figure 8). The appearance or loss of one species results in a dramatic shift in the index value.

As a composite measure, the Shannon-Wiener information index masks inter-species changes in abundance, and changes in the index are difficult if not impossible to relate back to community composition. For example, if one species has a disproportionate effect on another, their reciprocal changes in abundance will not be reflected by Shannon-Wiener. Clearly, one would hope to be able to track these changes in a community monitoring program. Finally, the Shannon-Wiener information index violates the assumptions of normal parametric statistics, and thus must be analyzed with less powerful non-parametric measures.

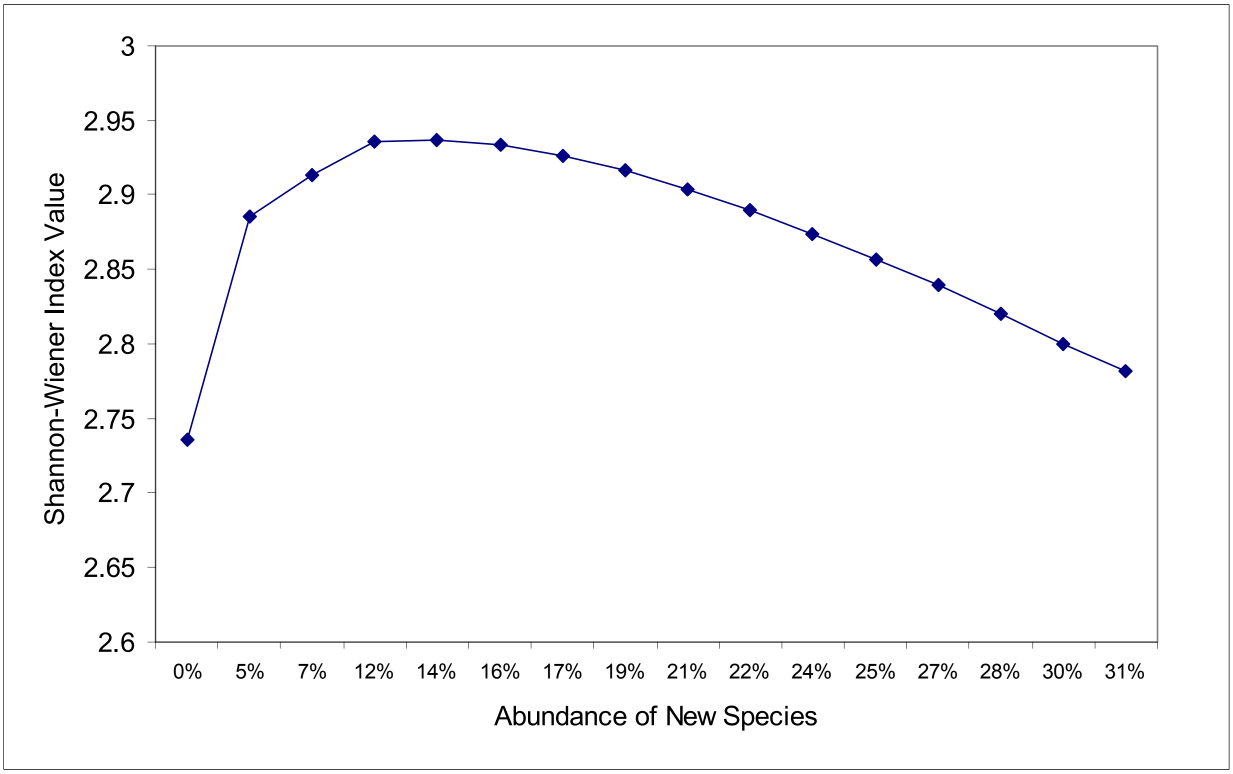

Figure 9. Shannon Wiener index change resulting from the appearance of a new species. This figure illustrates the effects of a new species on the Shannon-Wiener information index. As a new species is discovered, it dramatically reduces the value of the index, and then the index rebounds as that species becomes more abundant. Increased abundance then depresses the index again. The slope and break point of the relationship is determined by the original community richness and evenness.

A better choice for assessing relative changes within the entire community is the Bray-Curtis dissimilarity index (Provencher et al. 2002). This index accounts for each species' changes in abundance and is not disproportionately impacted by the addition or loss of species.

This index is calculated as:

$$1 - ((\text{Sum } |X_{ij} - X_{i(j+z)}|) / \text{Sum } (X_{ij} + X_{i(j+z)}))$$

where:
I = species
j = year
z = time increment

A similar measure of change over time is the squared Euclidean distance method of Krebs (1989 and Ratliff and Mori (1993). Either provides a robust, understandable measure of community change. An additional, and significant benefit of these indices is that they provide an easily understandable comparison between reference and other sites (Provencher et. al. 2002).

APPENDIX C. Sample data sheet for avian monitoring (Starts following page) . An Exel spreadsheet has also been provided to allow staff to customize this form.

Avian Monitoring

Site Name		**Date**	
Observation Point		**Time**	
***Wind Speed (0-5)**		**Observer**	
***Sky Code (0-7)**			
***Noise Code (0-7)**			
Air Temperature			

*see back of sheet for code descriptions

	0 - 3 minutes			**3 - 5 minutes**			**5 - 10 minutes**		
Species	**Males / Females / Imm.**			**Males / Females / Imm.**			**Males / Females / Imm.**		

Beaufort Wind Codes

0 Calm (<1mph) smoke rises vertically
1 Light Air (1-3 mph) smoke drifts, weather vane inactive
2 Light Breeze (4-7 mph) leaves rustle, can feel wind on face
3 Gentle Breeze (8-12 mph) leaves and twigs move around, small flags extend
4 Moderate Breeze (13-18 mph) moves thin branches, raises loose papers*
* Do not conduct survey at Level 4 or above
5 Fresh Breeze (19 mph or greater) small trees begin to sway

Sky Codes (note 3 and 6 are not valid code numbers)

0 Few Clouds
1 Partly cloudy (scattered) or variable sky
2 Cloudy or overcast
4 Fog or smoke
5 Drizzle or light rain (not affecting hearing ability)
7 Snow

Noise Codes

0 No appreciable effect (e.g., owl calling)
1 Slightly affecting sampling (e.g., distant traffic, dog barking, one car passing)
2 Moderately affecting sampling (e.g., nearby traffic, 2-5 cars passing)
3 Seriously affecting sampling (e.g., continuous traffic nearby, 6-10 cars passing)
4 Profoundly affecting sampling

Observer's Comments

APPENDIX D. Sample datasheet for collecting vegetation data. (starts following page).

Vegetation structure codes. The code is a five character string incorporating code for diameter (for forest and woodland stands) or height (for shrubland and grassland stands), canopy cover, and canopy layering (strata) (from Hall et al. 1995).

Tree stem size class	SA	sapling	20 trees per acre 1 - 4.9 inches dbh*
	PO	pole	15 trees per acre 5 - 8.9 inches dbh
	MT	medium tree	10 trees per acre 9 - 20.9 inches dbh
	LT	large tree	10 trees per acre 21 - 31.9 inches dbh
	GT	giant tree	5 trees per acre > 31.9 inches dbh

* This applies to the largest trees present. A class is determined by the average dbh of the number of trees per acre indicated.

Shrub/Grass height class:	HE	Herbland. Grasses and herbs are the only lifeform present.
	LS	Low shrub. Shrubs are 0 - 1.5 feet tall.
	Ma	Medium shrub. Shrubs are 1.6 - 2.5 feet tall.
	Mb	Medium tall shrub. Shrubs are 2.6 - 4.0 feet tall.
	Ta	Tall shrub. Shrubs are 4 - 6.5 feet tall.
	Tb	Very tall shrub. Shrubs are ≥ 6.5 (and < 16.5) feet tall.
Cover class:	Na	< 10 percent canopy cover.
	Oa	≥ 10 and < 15 percent canopy cover.
	Ob	≥ 15 and ≤ 25 percent canopy cover.
	Ma	> 25 and ≤ 40 percent canopy cover.
	Mb	> 40 and ≤ 66 percent canopy cover.
	Da	> 66 percent cover.
Strata	N	No strata.
	E	One stratum with < 30 percent difference in height.
	U	Two or more strata (of the same life form) with > 30 percent difference in height. If shrubland, a second shrub strata must have ≥ 25 percent cover. If herbland or grassland, a second herb or grass strata must have ≥ 10 percent cover (including cryptograms).

Tree stem size classes.

Code	Size class	Range
S1	seedling 1	< 6.0 inches tall
S2	seedling 2	> 6.0 inches
SA	sapling	1.0 - 4.9 inches dbh
PO	pole	5.0 - 8.9 inches dbh
MT	medium tree	9.0 - 20.9 inches dbh
LT	large tree	21.0 - 32.9 inches dbh
GT	very large tree	33.0 and greater

Tree stem decay classes.

Code	Status	Description
Standing dead		
SD1	decay class 1	bark, stemwood, and fine branch structure is intact
SD2	decay class 2	few limbs and no fine branches are present; the bark is partially broken; some stem decay may be present
SD3	decay class 3	only limb stubs are present; the bark is broken and sloughing; stem decay is evident
SD4	decay class 4	few limb stubs are present; the stem is usually broken and with evident decay; little bark remains
SD5	decay class 5	no limb stubs are present; the stem is broken and rotten; no bark remains
Dead and down		
DD1	decay class 1	bark, stemwood, and fine branch structure is intact
DD2	decay class 2	few limbs and no fine branches are present; the bark is partially broken; some stem decay may be present
DD3	decay class 3	only limb stubs are present; the bark is broken and sloughing; stem decay is evident
DD4	decay class 4	few limb stubs are present; the stem is usually broken with evident decay and conforming to microtopography; little bark remains
DD5	decay class 5	no limb stubs are present; the stem is broken, rotten and partially integrated into the soil; no bark remains

Vegetation Data

Line-point Intercept

Site Name

Date

Plot Number

Observer

Line Number

Recorder

page of

Point	Top Canopy Species	Lower Canopy Layers					Soil Surface
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							

STAND TREE STEM DATA

Plot ID: ______________________________ Date: ____________________________

Site: ____ Observers:__

Species	Status	Size Class	Age	Height	Tally

STAND TREE STEM DATA

Plot ID: ______________________________ Date: ____________________________

Site: ____ Observers:__

Species	Status	Size Class	Age	Height	Tally

COARSE-SCALE COMPOSITION AND STRUCTURE

SITE ______________________________ CELL ID _________________ DATE _________
Observer(s) ____

Transect # _______

Covertype	Plant Association	Structure	Dist

Transect # _______

Covertype	Plant Association	Structure	Dist

Transect # _______

Covertype	Plant Association	Structure	Dist

Transect # _______

Covertype	Plant Association	Structure	Dist

Transect # _______

Covertype	Plant Association	Structure	Dist

Transect # _______

Covertype	Plant Association	Structure	Dist

APPENDIX E. Ninety percent Binomial Confidence intervals for all possible frequencies when 240 points are sampled.

	Confidence Interval	
Freq/Cover	Lower	Upper
1.00	0.148	2.600
2.00	0.824	4.330
3.00	1.377	5.408
4.00	2.278	6.965
5.00	2.910	7.975
6.00	3.561	8.969
7.00	4.563	10.436
8.00	5.246	11.400
9.00	6.287	12.831
10.00	6.992	13.775
11.00	7.703	14.713
12.00	8.781	16.108
13.00	9.506	17.032
14.00	10.603	18.409
15.00	11.340	19.322
16.00	12.082	20.231
17.00	13.200	21.588
18.00	13.950	22.488
19.00	15.082	23.832
20.00	15.840	24.724
21.00	16.601	25.614
22.00	17.747	26.943
23.00	18.515	27.826
24.00	19.671	29.147
25.00	20.444	30.024
26.00	21.220	30.899
27.00	22.388	32.207
28.00	23.170	33.077
29.00	24.345	34.377
30.00	25.131	35.242
31.00	25.920	36.105
32.00	27.105	37.396
33.00	27.898	38.254
34.00	29.090	39.538
35.00	29.888	40.392
36.00	30.686	41.244
37.00	31.888	42.520
38.00	32.690	43.368
39.00	33.898	44.637
40.00	34.705	45.481
41.00	35.513	46.324
42.00	36.729	47.585
43.00	37.541	48.424
44.00	38.762	49.680
45.00	39.578	50.515
46.00	40.396	51.349
47.00	41.625	52.596
48.00	42.446	53.426
49.00	43.680	54.668
50.00	44.505	55.495
51.00	45.332	56.320
52.00	46.574	57.554
53.00	47.404	58.375
54.00	48.651	59.604
55.00	49.485	60.422
56.00	50.320	61.238
57.00	51.576	62.459
58.00	52.415	63.271
59.00	53.676	64.487
60.00	54.519	65.295
61.00	55.363	66.102
62.00	56.632	67.310
63.00	57.480	68.112
64.00	58.756	69.314
65.00	59.608	70.112
66.00	60.462	70.910
67.00	61.746	72.102
68.00	62.604	72.895
69.00	63.895	74.080
70.00	64.758	74.869
71.00	65.623	75.655
72.00	66.923	76.830
73.00	67.793	77.612
74.00	69.101	78.780
75.00	69.976	79.556
76.00	70.853	80.329
77.00	72.174	81.485
78.00	73.057	82.253
79.00	74.386	83.399
80.00	75.276	84.160
81.00	76.168	84.918
82.00	77.512	86.050
83.00	78.412	86.800
84.00	79.769	87.918
85.00	80.678	88.660
86.00	81.591	89.397
87.00	82.968	90.494
88.00	83.892	91.219
89.00	85.287	92.297
90.00	86.225	93.008
91.00	87.169	93.713

92.00	88.600	94.754
93.00	89.564	95.437
94.00	91.031	96.439
95.00	92.025	97.090
96.00	93.035	97.722

97.00	94.592	98.623
98.00	95.670	99.176
99.00	97.400	99.852
100.00	98.760	100.000