Statistical Considerations in Monitoring Birds Over Large Areas

Douglas H. Johnson

Abstract—The proper design of a monitoring effort depends primarily on the objectives desired, constrained by the resources available to conduct the work. Typically, managers have numerous objectives, such as determining abundance of the species, detecting changes in population size, evaluating responses to management activities, and assessing habitat associations. A design that is optimal for one objective will likely not be optimal for others. Careful consideration of the importance of the competing objectives may lead to a design that adequately addresses the priority concerns, although it may not be optimal for any individual objective. Poor design or inadequate sample sizes may result in such weak conclusions that the effort is wasted. Statistical expertise can be used at several stages, such as estimating power of certain hypothesis tests, but is perhaps most useful in fundamental considerations of describing objectives and designing sampling plans.

This paper presents a personal view on the topic of monitoring birds over a large area, such as a national park, forest, or wildlife refuge. My ideas derive from several years experience as a consulting statistician in a natural resource management agency. In recent years, especially, I have been asked by managers for help in establishing surveys for nongame birds. Most of my experience is with U.S. Fish and Wildlife Service managers, who are responsible for fairly large areas such as national wildlife refuges, as well as groups of smaller units such as waterfowl production areas. My limited interactions with land managers in other agencies suggest that their concerns are essentially similar. My emphasis here is on the statistical (especially the design) aspects of monitoring, rather than on the field methodology. Thus, I will not concentrate on questions such as whether point counts should be used in lieu of line transects. Useful references about field methodology include Ralph and Scott (1981), Verner (1985), and Ralph and others (1993) for birds; Lancia and others (1994) discussed methods of population estimation for wildlife more generally.

When establishing a monitoring program, many questions arise: What should be done? When should it be done? How should it be conducted? Who should do it? Where should sites be located? Why should it be carried out? And, especially for administrators, how much will it cost?

The major point of my paper is that the "Why" question is overriding and should be clearly addressed first. Understanding—and agreeing to—the objectives of a monitoring program will facilitate addressing the remaining issues. Knowing why something is to be done will dictate answers to the other questions (Roberts 1991; Usher 1991). As Green (1984:298) said, "environmental biologists tend to worry too much about the statistical methodology and too little about the question."

When asked by managers to help design a monitoring scheme for birds, I often become a reflective listener, but aggressively so. I first ask them why they want to monitor birds, and when they tell me, I ask more questions. If they say they want to know what is happening to bird populations on the area they manage, I ask why they want to know, and what is the area of concern, and do they care what is happening elsewhere. If they say they want to know what effect their management actions are having, I ask them why, and which practices are of concern, and how much land they affect. While such inquisitions can prove irritating to the managers, and remind them of youngsters who respond to every answered question with another "Why?", I have found that the process does tend to refine the manager's thinking. He or she will eventually settle on one or more specific needs that should be addressed. And that usually leads to answers to the other questions.

In several recent workshops with U.S. Fish and Wildlife Service land managers, I asked them why they wanted to monitor birds, in this case nongame birds. Responses ran the gamut. A number of managers wanted to contribute to broad-scale, coordinated monitoring and research efforts. They expressed an interest in conducting North American Breeding Bird Survey (Peterjohn 1994) routes, or supporting MAPS (Monitoring Avian Productivity and Survivorship) (DeSante and others 1993b) stations, or contributing to the U.S. Forest Service's bird monitoring plan (Manley 1993). Others more generally wanted to provide better information on distribution of birds or their habitat affinities, or to identify threats to their viability.

Some managers indicated an interest in local-scale issues. Managers wanted to assess the value of their area to nongame birds. Others wanted to determine when and how their area is used by different species. Some U.S. Forest Service managers indicated that monitoring of nongame birds is required by laws or regulations.

Several managers' interests were tied to the general "health" of their managed areas. Some managers wanted to monitor nongame birds to indicate habitat quality and environmental health. (See Temple and Wiens [1989] for an interesting discussion of using birds as bio-indicators.) Others mentioned monitoring indicator species. Several managers viewed nongame birds as part of the ecosystem approach, recently adopted by the U.S. Fish and Wildlife Service and other federal agencies (Malone 1995).

Beyond the objectives listed above, three reasons were most consistently mentioned for monitoring in managed

In: Bonney, Rick; Pashley, David N.; Cooper, Robert J.; Niles, Larry, eds. 2000. Strategies for bird conservation: The Partners in Flight planning process; Proceedings of the 3rd Partners in Flight Workshop; 1995 October 1-5; Cape May, NJ. Proceedings RMRS-P-16. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Douglas H. Johnson, Northern Prairie Wildlife Research Center, USGS Biological Resources Division, 8711 37th Street SE, Jamestown, ND 58401.

areas, and these will be the focus of the remainder of this paper. These were (1) to get presence/absence information for species, (2) to monitor population trends, and (3) to evaluate the effects of specific management actions. I tried to demonstrate to the managers that the optimal monitoring design will vary, depending on which objective is paramount. As we treated each objective, we discussed various methods available to meet the objective and evaluated the advantages and disadvantages of each method.

Obtaining Presence/Absence Information

If the primary objective is to learn which species use the area, possibly by season, a variety of methods can be employed. Point counts (Ralph and others 1995b), transects (Mikol 1980; Seber 1982), territory mapping (Hall 1964; Val Velzen 1972), nest searching, and other systematic methods will provide information about species using the area. But these methods are labor-intensive relative to the information obtained.

A more suitable method is to conduct aggressive searching of all habitats on the site. Mist-netting (Karr 1981) might be used to augment the database. Nocturnal surveys will improve the chance of detecting owls, rails, goatsuckers, and certain other species that are more active at night. Playing recorded calls and listening for responses is valuable for certain secretive species (Johnson and others 1981). Adding observations of species detected incidental to other activities also will be useful. Managers should develop a system for staff and visitors to record such sightings, verify them, and add them to the data base. Also, there may be historical data that can be included, although old occurrence records may tell little about current presence.

Addressing this objective lends itself to a cooperative effort with the local birding community, either individuals or organized clubs. The best information on the bird species using an area can be obtained by encouraging active and proficient birders to cover the area. Managers might make it a contest to see how many species can be found. Birders should visit all habitat types, as long as they do no damage and do not interfere with management activities, and visiting birders should be encouraged to share their sightings.

"Big Days," on which birders attempt to find as many species as possible during a single 24-hour period, can be encouraged, with the proviso that sightings need to be restricted to the area of concern. I suggest that such activities, conducted during the various seasons for several years, would provide an excellent list of species using the site. That list could then be augmented by incidental observations and species found during other surveys.

This objective ties in nicely with atlas projects, which are intended to determine the breeding distribution and reproductive status of breeding birds. These are usually done by state or province, and many states have recently completed or are currently conducting atlas projects. Smith (1990) discussed techniques for atlas projects.

A key consideration is that records of birds need to be verified, especially for unlikely species. One advantage of using proficient birdwatchers is that they are better able to recognize the species. Many state birding societies have developed systems to evaluate unusual sightings. Photographs or other evidence may be required in certain circumstances. I suggest using those systems to verify species records for managed sites. Also, some thought should be given to indicating the relative abundance of, or likelihood of a bird watcher finding, each species.

A final word about this objective: when pressed, managers generally acknowledge that learning which species use their site is *not* their foremost objective. As more effort is devoted to meeting the objective, the longer this list of species will become. Eventually any species added will be a rarity, especially one whose normal range is distant from the site. The value of a list of extralimital species is questionable. Although managers—and visiting birders—like to know which species are likely to be found on a site during a particular season, managers usually need information addressing the other two major objectives.

Monitoring Bird Population Changes

Many managers indicated that they wanted to know how bird populations on their areas were changing over time. For that objective, the *ad hoc* procedures recommended above are not very suitable. General references on ecological monitoring include Goldsmith (1991) and Spellerberg (1991). One of the most important considerations for monitoring is that the system needs to be repeatable. Except in very unusual situations, results of a monitoring effort in any year will provide but an index to the population size of each species. It is useful to have the index close to the actual population size (that is, to monitor a large fraction of the population), but it is far more important to have the index strongly correlated with the population size (Johnson 1995).

Some guidelines will help ensure that an index is accurate. The advice is simple to state but not always easy to follow. Repeatability demands that standardized methods be used. Protocols must be written down, understood by the participants beforehand, and adhered to closely. Observers vary in their abilities (Faanes and Bystrak 1981; Sauer and others 1994b). For this reason, it is best to use the same observer for as many years as possible. If it is necessary to change observers, careful training will minimize the disruption in the monitoring scheme (Kepler and Scott 1981; Hanowski and Niemi 1995).

For monitoring purposes, it is probably best to use the same locations for counts in all years. A system with some old and some new locations each year might offer somewhat better statistical properties, but I doubt that it is worth the effort. Observations should be made in all major habitats. As a rough guide to sample sizes, I would place sampling locations in each habitat in proportion to the area of that habitat on the site, times the (expected) density of birds in the habitat. Simply stated, put more effort in more common habitats or habitats with a lot of birds. Also, areas of special importance to birds (e.g., breeding colonies, special habitats) should be monitored, even if their areal extent is small.

I have not discussed the actual bird survey methods. Territory mapping can be used, but again is labor-intensive and restricted to small areas. Mist-netting may be useful in certain habitats, but I would suggest point counts or line transects as methods that detect large numbers of birds and, most importantly, are repeatable.

The point count method, the current favorite among bird survey methods, has a number of advantages over other techniques (Ralph and others 1995b). It does not, however, provide estimates of abundance or density. Density estimates can be made if distance sampling is used (Buckland and others 1993), or if a fixed-radius circle is employed in the point count such that all birds within that radius are detected and none beyond the radius are included.

A question often arises about the use of roadsides as routes for point counts or other survey methods. Certainly, if one is interested in getting an accurate count of each species on the site, those species that favor roadsides will be detected more often from roadside counts than their actual abundance would indicate (Rotenberry and Knick 1995). The opposite would be true, of course, for any birds that avoid roads (although Keller and Fuller [1995] found no such pattern in second-growth forest). Is this a problem if monitoring change is the objective? No, not if detectability remains more-or-less constant from year to year. A systematic roadside survey may, as an example, record 10 percent of all Mourning Doves (*Zenaida macroura*)—which favor roadsides—on the area, but only 3 percent of all Baird's Sparrows (*Ammodramus bairdii*)—which do not. So using survey results to compare numbers of Mourning Doves and Baird's Sparrows would be misleading. But, as long as those percentages remain roughly the same from one year to the next, the survey will adequately portray changes in the population size of either species.

Offsetting the potential bias of roads is their convenience. Routes that follow roads offer logistical advantages. During the same amount of time, one might be able to conduct only a third as many counts off-road than along a road, and even fewer if the terrain is difficult. Moreover, the extent of the roadside bias will depend on the kind of road; a narrow trail winding through a habitat will pose less of a problem than a freeway (Hutto and others 1995).

One remedy is to conduct some on-road counts and some off-road counts in similar habitat, and compare both the numbers of birds detected and the amount of time needed to carry out the surveys. Analysis of resulting data should be informative. My suspicion is that, in most cases, the logistic convenience will more than outweigh any roadside bias. If important habitats do not have roads or trails through them, then the sampling design needs to be augmented to incorporate those habitats. Also, routes along roads would not be feasible in smaller areas.

A final thought on monitoring: To be of most value, monitoring should be conducted as part of a larger effort, so that trends on the site can be compared with those off-site. Related to this point is my opinion that monitoring bird populations may not be sensitive to effects of management actions taken on the site (Roberts 1991). Monitoring is useful when the questions to be asked cannot be anticipated (Hamel and others, this proceedings). If responses to known actions are desired, alternatives preferable to general monitoring are available (Cooper and others, this proceedings).

Evaluating Effects of Management Actions

Managers manage. Whether by their own direct actions, the activities of cooperators, or passive inaction, managers cause changes on the lands for which they are responsible. They want and need to know what effects they are having on target as well as nontarget species. This objective was rated as most important by the managers I talked with. Some suggestions follow, directed toward assessing the effects of management actions on nongame birds (which historically have been nontarget species for most managers).

First, a general monitoring scheme is unlikely to detect changes in bird populations due to management actions unless either the monitoring program is very intensive, the actions have very marked consequences, or a lucky event has occurred. Most often, management practices—such as prescribed burning, grazing, or cutting timber—affect only part of the area during any particular year. Unless the bulk of a species are in the affected area, only a fraction, say 15%, of the birds may be affected. And fluctuations of that magnitude are common in most bird populations and well below the level likely to be detected under routine monitoring.

To identify effects of a management action, it is better to focus on the species and habitats likely to be affected by the action. (Also, consider how the treatment might affect the species: density, species composition, productivity.) Sampling should be heavy in areas to be treated and in similar areas that will not be treated (control, or comparison, areas). Samples should be taken before the treatment, and for several years following. Sites to be treated and sites to be used for comparison should be selected randomly from a set of similar sites. A number of independent treatments should be applied, affording some replication. These three features—treatments and controls, randomization, and replication—form the basis of sound experimental design (Fisher 1947).

It may be useful to pause here and contrast three scientific approaches: experimental, observational, and sample survey. In an experimental approach (e.g., Cochran and Cox 1957; Milliken and Johnson 1984), one is interested in the effect of some treatment, such as grazing, on some variable, such as the density of a particular bird species. The process is manipulative: the unit is treated or perturbed by the experimenter. He or she then can select control sites, randomize the treatment and control, and replicate. In an observational study (Cochran 1983), one likewise is interested in a treatment effect, but the process is passive: the treatment is applied by someone else (or nature) and control sites may not be available, randomization is not feasible, and replication may not be possible. In a sample survey (e.g., Cochran 1977; Williams 1978), interest lies not in a treatment effect but in some population feature, for example, the density of Baird's sparrows in grazed fields in North Dakota. Hurlburt (1984) termed such surveys "mensurative studies." Like an observational study, the process involved in a sample survey is passive.

The experimental approach is stronger than an observational approach for understanding effects of a treatment (Skalski and Robson 1992, a book that is useful more generally than for mark-recapture studies, as its title would suggest). Experiments are, unfortunately, more demanding and frequently impractical or even impossible. In such situations, the observational approach and some other techniques offer partial remedies. An observational study arises largely because one cannot assign treatments and controls. One useful tool for resulting data is matching, trying to find a unit that is as similar as possible to the treated unit except for the application of the treatment. Another tool is statistical control, essentially trying to control for other variables during the analysis, rather than during the design. Instead of randomly assigning treatments and controls, one measures a number of other variables—such as soil type, precipitation, and vegetation, in our example—and attempts to statistically adjust for their effects. Cochran (1983) is a valuable reference for this approach.

Sometimes a treatment will be applied, but the investigator has no hand in where it will be done; randomization is impossible. Perhaps, as an example, a timber sale will result in major tree-removal in a certain area, and that area is predetermined. Suppose there is interest in understanding the effects on bird populations. A BACI (Before and After, Control and Impact) design may be appropriate (Skalski and McKenzie 1982, Stewart-Oaten and others 1986). The idea here is to measure bird populations before and after the timber cutting, and in similar (control) areas as well as in the treated (impact) area. The methodology was developed in the ecological literature for situations such as power plant sitings, and has been described by Eberhardt (1976), Green (1979), and Stewart-Oaten and others (1986), among others.

A related method is intervention analysis. In that approach, some event has occurred (such as banning DDT) and interest lies in detecting if a change in some variable (such as the number of Bald Eagles *[Haliaeetus leucocephalus]* in an area) occurred coincidentally with the event. Intervention analysis uses time-series methods to determine if such a change took place (Box and Tiao 1975). It is used when the investigator has no control over the event and the study generally is designed after the event, rather than before (Eberhardt and Thomas 1991). The method requires a rather long series of measurements both before and after the event, which limits its applicability. BACI, intervention analysis, and related methods require assumptions and analyses based on models to separate true effects from confounding influences of space and time (Skalski and Robson 1992:162). As such, they are inferior to manipulative experiments.

Randomization is important in an experimental setting in the context of assigning treatments to sample units. That aspect of randomization is far more important than another way randomization is used: to select points at which a response is measured. For example, suppose one is interested in the immediate effects of burning a grassland on densities of Sprague's Pipits (*Anthus spragueii*). Ideally, one should have a number, say 20, of similar fields and randomly choose 10 of those fields to burn. The remaining 10 fields are left unburned to serve as controls, or as a basis of comparison to the burned fields. One (or more) point count is made in each field. This approach is how randomization (and replication) should be used, and provides a sample size of 10 in each of the two groups. The error degrees of freedom for a t test would be $20 – 2 = 18$, and would likely provide some strong inferences about the treatment effect.

Contrast this to the situation in which one field is burned and one left idle, but that 10 point counts of Sprague's Pipits are made in each field after the treatment. The points are selected randomly within each field. Here randomness may help us get a good estimate of Sprague's Pipits on each field (or it may not; a systematic sample would probably be preferable), but it does nothing to ensure that the field selected for treatment did not differ in some important way from the field selected for the control; a random selection of fields to which to apply the treatment would have done that. Also, this randomization does not increase the sample size for the comparison of burned versus unburned fields; there is one sample per treatment, the error degrees of freedom are $2 – 2 = 0$, and a t test (or any other test) cannot validly be conducted. Treating the 10 point counts done within each field as sample units would be a fine example of pseudoreplication (Eberhardt 1976; Hurlbert 1984). Eberhardt and Thomas (1991) also noted the common error of confusing control of treatments with control of the observational process.

Occasionally the other essential component of the experimental approach, replication, is not achieved. (Sometimes it is impossible, but too often it is simply overlooked.) Suppose the densities of Baird's Sparrows are 10 per 100 ha in a grazed field and 20 per 100 ha in an ungrazed field. What does that tell us about the effect of grazing on Baird's Sparrows? Almost nothing; the fields may differ in other important aspects (the probability of which could be reduced by randomization, or the effects adjusted for by careful analysis of measured covariates). And, equally importantly, the difference may be due to chance; had two other identical fields been considered, densities might have been 20 per 100 ha in the grazed field and 10 per 100 ha in the ungrazed one.

This possibility is one reason why replication is so important; it tends to average out the effects of random events, so that treatment effects stand out. If replication at a single site is impossible, meta-analysis may—and I emphasize may—offer a potential remedy. Meta-analysis was designed to reach an understanding about the value of some parameter (in our case, treatment effect) from numerous studies in which the parameter has been estimated. Each study produces an estimate of the parameter, along with a measure of its uncertainty, such as standard error. Meta-analysis basically is an analysis of analyses, in which the individual studies form sample units for the larger study. The methodology was not developed for situations in which the individual studies themselves were not replicated and did not estimate standard errors, but perhaps it could be adapted for such situations. For example, if densities of Baird's Sparrows on grazed and ungrazed fields were measured for a number of sites in the mixed-grass prairie, some general conclusions might be reached, even if no single site replicated the comparison sufficiently to produce reliable answers to the question. One danger is that, at each site where grazed and ungrazed fields were compared, the field where grazing had been applied might differ in some fundamental way from the ungrazed field. This is a problem that randomization is designed to alleviate; meta-analysis can address only the lack of replication. Some references on meta-analysis are Hedges and Olkin (1985), Gurevitch and Hedges (1993), and Arnqvist and Wooster (1995).

Replication within a study is important, but replication of studies in space and time is far more important (Hawkins 1986). To know that the density of Baird's Sparrows was higher in an ungrazed field than in a grazed field at one site in one year is useful, perhaps. Knowing that the result is general, that it holds for mixed-grass prairies in general, is far more valuable. This more universal conclusion can be reached only by replicating the study in areas throughout the mixed-grass prairie and in multiple years. Quoting Eberhardt and Thomas (1991:57), "Since truly definitive single experiments are very rare in any field of endeavor, progress is actually made through sequences of investigations."

Compromise Designs

Typically, managers, like the rest of us, want it all. They want to know what species occur on their areas, they want to monitor population changes, and they want to know the consequences of their management activities. As described above, the optimal design for any single objective will be suboptimal, and perhaps nearly worthless, for another objective. Can a compromise design be developed that will allow all objectives to be met to some degree, even if not optimally?

I suggest the answer is yes, and offer some thoughts. The core of such a design would focus on monitoring, with incremental work to address more directly the other two objectives. Suppose a general monitoring scheme is in place, following guidelines mentioned above. It might be a series of point counts, mostly along secondary roads or trails, possibly with some off-road points for comparison. All important habitats (importance due to either areal extent or use by birds) should be covered. Most critically, the monitoring protocols should be clearly described and followed closely in each year. This plan would directly address the monitoring objective.

For the checklist objective, results of the monitoring can be supplemented with incidental information gathered during other activities and with records provided by visiting birdwatchers. Extra efforts such as mist-netting, playbacks, and nocturnal surveys may add species not otherwise detected. And cooperating with local or state birding groups to use a system of verifying records and to develop checklists will likely prove worthwhile.

To assess the changes in bird populations due to management actions, I suspect that the routine monitoring scheme by itself will prove inadequate. But results of that scheme can form the basis of a useful evaluation. I would suggest augmenting the monitoring design by adding more sampling points in areas to be affected by the management actions and in areas not to be affected. The basic criteria of experimental design should be met if at all possible: control and treated sites, random assignments of treatments, and replication. So if a treatment is planned for a particular site (preferably, the site had been randomly chosen for the treatment, and it is only one of several replicates), the number of sampling points at that site and at others similar to it (comparison sites) should be increased. Those sampling points should be surveyed both before the treatment is applied and for several years after, depending on how long the treatment is likely to have an effect. Results of the basic monitoring scheme may prove useful (1) for selecting sites that are relatively similar, to use as potential treatment and control sites in the experiment; (2) for understanding the variation in counts so that necessary sample sizes can better be determined; and (3) as part of the sample used to estimate the effect of the treatment.

Conclusions

At the early stages of learning which species are using an area, developing a checklist of species present is a valuable endeavor. An intensive effort by cooperating birdwatchers, augmented by recording of incidental observations, can produce the desired list in relatively short order. A system for verifying sightings of unusual species should be adopted.

For long-term monitoring, point counts or similar methods are applicable. They should be conducted in all important habitats. I suspect that roadsides can be used for routes, if they are appropriately distributed, but any biases can be evaluated in a comparative study. The most critical concern is that the monitoring be systematic and repeatable; standardized protocols, strictly adhered to, and a training program for observers will help to ensure the quality of the program.

If the effects of management on bird populations are to be evaluated, the three foundations of experimentation should be assured: treatment and control sites, randomization of treatment, and replication. Emphasis should be on areas to be treated and on similar areas left untreated. Some partial remedies, briefly mentioned above, may be helpful if one or another cannot be met.

Managers should also consider the importance of habitats on their areas, and the species for which their areas are important. The manager of a refuge containing unusual fen wetland habitat, or a small stand of old-growth trees, should give special emphasis in a monitoring plan to bird species that rely on those habitats. And a manager of a refuge in the mixed-grass prairie should ensure that bird species that require such grasslands, such as the Baird's Sparrow, are monitored before concerning him- or herself with widespread and more common species that do not need those habitats, such as Brown Thrashers (*Toxostoma rufum*), for example. In that sense, the monitoring plan should mirror the management priorities.

Acknowledgments

I am grateful to the many wildlife managers with whom I have consulted, and who helped me learn as I was trying to teach. This paper benefited from reviews by L. D. Igl, W. E. Newton, J. R. Sauer, and T. L. Shaffer.