Studies in Avian Biology No. 6:570–574, 1981.

ADAPTING GENERALIZED INSTRUCTIONS TO SPECIFIC SITUATIONS IN PLANNING COUNT WORK

A. J. ERSKINE[1]

ABSTRACT.—The decision to carry out a bird count leads to many other decisions, many of which lie well outside the interest or expertise of the would-be counter. The intensity of effort required scares off many potential candidates, but the amount required is debated even by experts. The need for precisely defined and measured areas and for detailed habitat descriptions are further stumbling blocks, upon which specialists disagree. Even the methods to be used are open to debate, depending on the purposes to which they are to be put. Many criticisms levelled at bird count work stem from misconceptions of objective, scale, or perspective, and no single set of guidelines will satisfy all needs.

This conference is about counting birds. Counting obviously means different things to different people, but the common denominator is "a systematic effort to count birds," for any of a wide variety of reasons. Within that general definition, bird-counting exercises may be as wide ranging as the cooperative Breeding Bird Survey (Bystrak 1981) or as confined as a study of a single plot, as rigidly standardized as the "Mettnau-Reit-Illmitz" migration monitoring program (Berthold and Schlenker 1975), or as unstructured as the Christmas Bird Counts (Arbib 1981). These diverse understandings of counting reflect an equally wide array of reasons for counting birds, among which may be mentioned the study of ecological relationships, the monitoring of trends relative to ongoing environmental changes, measuring impacts of land-use or pollution, management of directly exploited birds or pest species, or recreation. The last is by no means the least, since most of us who work in ornithology do so because we enjoy it, whether or not we make a living from it.

What sets counters apart from many bird watchers is their attempt to count birds systematically. All systems impose constraints, which in their turn call for decisions. Many of the decisions required of counters seem to have little to do with birds or our interest in them, since they revolve around the objectives, the methods, the results, or the interpretation of them. One of the most basic decisions is whether to pursue absolute numbers or densities or merely to obtain relative indices to populations. No satisfactory decision can be reached on this point until one has considered the resources of time, personnel, and equipment available to the counter, as well as the objectives in undertaking the census. Bird counting has come a long way in the 35 years since Kendeigh's (1944) review, but most counters still start out virtually in isolation, making their own decisions—and mistakes—the first time around.

[1] Canadian Wildlife Service, P.O. Box 1590, Sackville, New Brunswick E0A 3C0, Canada.

The purpose of this presentation is to explore some of the basic questions and constraints facing counters, and to outline what I believe to be the best types of decisions in dealing with them. Obviously this will be a biased viewpoint; I have never used transect counts systematically, and it will be obvious that I see mapping censuses of measured plots (Williams 1936, Enemar 1959) as fundamental to many censusing efforts. Furthermore, my remarks are focussed on censusing more or less sedentary populations, and not all my generalizations will fit counts of migrating birds.

BASIC QUESTIONS IN COUNTING

I referred already to the basic division of counting into absolute vs. relative counts. Except with a few very scarce and/or localized species, such as Kirtland's Warbler (*Dendroica kirtlandi*) and North Atlantic Gannet (*Morus bassanus*), whose entire breeding populations have been counted more or less directly, all censusing involves the counting of samples from a population. In absolute census counts, the samples are drawn from a defined area of (supposedly) known extent; relative counts are assumed to have sampled the same area each time, without the size of the area necessarily being known. If one method gives truly absolute results, these *should* be comparable with results from other absolute methods; unfortunately, the truism that "all things are relative" applies also to "absolute" census methods, to a greater or lesser extent. Results obtained by relative methods, however, can only be compared with others obtained by precisely the same methods and usually by the same observer. Migration counts in particular describe only the situation at one moment in time, so are always relative. Thus, one of the basic constraints here is comparability. One must decide if one's results need readily to be compared with those of other people, or if comparisons are to be made only among one's own samples.

Another basic conflict in counting is between standardization and practicality. The fact that so many methods have been discussed in this conference suggests that no one method readily meets all possible needs. One may ask: is it really possible to restrict counting to a few standardized methods? And, can any method or methods be flexible enough to accommodate the various objectives and situations without being so general as to defy meaningful comparison of results? One easy answer, previously arrived at by some people (e.g., Berthold 1976), is "yes!" to both questions. To such people, all that is needed is to settle on a few rigorously standardized and very intensive methods, and to insist on them to the exclusion of all others. The data so obtained will be comparable, and of irreproachable accuracy; but, unfortunately, most areas will remain unsampled, since most counters, whether professional or amateur, will not be able to put in the required intensity of effort, even if they are willing to accept the prescribed methods as the only satisfactory ones. There must be some compromise between scientific rigor and practical usefulness, but the range of acceptable compromise is not very wide.

Specialized knowledge is another stumbling block to many would-be bird counters. Ability to identify birds is obviously a prerequisite to any serious count effort, and some methods demand the identification of every "tweet," "cheep," "chip," or half-heard song as well as of those birds that happen to stray into view. Skills of bird identification by sound have spread widely since the Breeding Bird Survey started in 1965–68 (Robbins and Van Velzen 1967), and few bird-watchers now are unwilling to accept such challenges, within the constraints imposed by hearing and sight. Descriptions of habitats pose a much greater burden to most amateurs, and some professional biologists are unfamiliar with many of the more common plant species in areas where they have studied birds for years. Particularly in cooperative projects involving amateurs, the most successful are those that demand of the cooperators only to count birds. Rules governing when and where to count them are not nearly as much of a problem as are directions to perform other activities that do not interest most volunteer observers.

Statistics pose additional problems in censusing. Recent preoccupation with computers and complex statistical treatments has tended to focus on methods giving data that can be mechanically converted into coefficients that only a specialist can interpret. Most of this is not essential, and much of it may be counter-productive if it overemphasizes use of a large number of small and inherently highly variable samples merely because other methods give results less suitable for computers. The idea that count results without attached confidence limits are intrinsically inferior to those with them, regardless of the methods used, is sufficiently absurd to a dispassionate view that it should not be allowed to sway decisions on censusing.

EXAMPLES OF CONSTRAINTS REQUIRING DECISIONS

The examples that follow are hypothetical but all based in part on real situations. Amateurs usually start out by taking part in cooperative efforts like the Christmas Bird Count, and as they gain expertise some move on to individual projects. Because these are spare-time undertakings, their scope is limited to early mornings, evenings, weekends, or holidays. Repeated transect counts to monitor migration within a year or summer or winter numbers between years may satisfy some lone observers for a while (e.g., Erskine 1968). These relatively low-intensity methods are usually inherently variable, so the results tend to be difficult to compare with those of others. Few such projects are continued for long enough that the observer can usefully compare his (or her) own results accumulated over time, and a common result is a mass of summarized but unanalysed data. (The same thing has also resulted from some transect surveys by professionals . . .). A probably more useful exercise for a lone amateur is a mapping census of a plot, but here too the neophyte encounters problems. Without an assistant, the plot will probably be paced rather than measured, which in many habitats will be less accurate. The censuser may know nothing of vinyl flagging tape, and end up with inadequately marked grid lines, to the detriment both of the results and of his enjoyment. Except in southern and far western parts of North America, the usable census period is only 6 weeks or even less (cf. Erskine 1976b), so fitting the required eight counts into that period calls for more than one count each week, whatever the weather. Usually, this forces the use of a study area close to home, and small enough that it can be censused on weekdays outside of working hours, unless the censuser is eager enough to devote nearly every possible weekend morning through that period to the census. And as already noted, many amateurs give up when faced with the habitat description for the plot; if they cannot recruit a botanically minded friend at this stage, the census may never be written up and results lost. For some amateurs, the first mapping census attempt is also the last, at least partly for lack of quite elementary instructions to assist in decisions on matters other than censusing birds.

A second example is drawn from the field of graduate studies. A student wishing to explore, say, relationships of habitat and bird communities sets up census plots and conducts mapping censuses throughout the year. However, territories can only be mapped in breeding season, as birds tend not to be stationary nor to advertise their locations at other seasons. Accordingly, so as to have a common numerical basis throughout the year, the results are worked up only as mean numbers of each species per count in each season, which will be comparable among themselves, between seasons, or between years. Such results are amenable to statistical comparisons, but they have only limited comparability with those from other studies, as the density indices so obtained are substantially lower than the absolute values to be anticipated. A decision not to estimate territories of breeding birds may have been made merely because territories were thought too subjective to permit statistical analysis, as well as because they applied only to the breeding season. However, unless the field data are placed in a permanent repository, no one else will be able to reconstruct the (more or less) absolute density figures that could have been derived from such mapping censuses; and no one will be able to do so from the same familiarity with the areas as would the censuser. Often even the thesis omits the absolute density figures, and the subsequent publication—if and when it emerges—almost invariably lacks this basic information, because the censusing was looked on only as a means to the student's own ends. Students directly concerned with methodology are perhaps more likely to publish the actual census results, especially if these involve innovations; but they and their supervisors need to remember that the comparative data they obtained by established methods may turn out to be more useful to others than the innovations that justified their study.

Next we may turn to a consultant with, say, a contract to monitor the effects on birds of a spray program against forest insect pests. Spraying is timed to a particular stage of the emergence of insects or their larvae, usually sometime in the middle of the birds' breeding season. Count methods chosen have to give data that can be compared from one part of the season to another, and have to involve samples taken over sufficiently wide areas as to average out the inevitable unevenness in application of sprays. Use of a few large plots risks some being missed altogether or else overdosed, while having many small plots involves so much edge effect that the results may be nearly impossible to interpret; transects often seem the only solution. The results are frequently highly variable even when all counts are conducted by the same observer, and some changes that are obvious to the observer in the field may not show up in the results because of the swamping effect (nearby songs drown out distant ones at high but not lower densities). The end results are seldom satisfying either to the censuser or to the agency employing him, as only acute effects are documented to any conclusive extent. Most decisions on methodology for short-term monitoring have been made on an ad hoc basis, or empirically, in comparison to what worked or didn't work last time. And many people who have tried to monitor forest spray programs have turned to other work in frustration. Consultants, of course, are in that game to make a living; any extra expenditure on a project reduces their potential profits. When they set out to conduct, say, an inventory or monitoring of a bird population, they will do so with as few surveys and as poorly-paid staff as they can get away with, expecially if they can pull political strings to ensure that rival firms with higher competence and/or standards cannot compete for the contract. In one such case, a contract was awarded to a consulting firm, who sublet it to a graduate student, who passed the bulk of the actual field work to a "birding bum," who ended up abandoning the job—and departing without telling anyone—at the height of the breeding season; there was a gap of 2 weeks before another sub-sub-contractor could be found and put to work.

Lest I be accused of pointing a finger only at outside groups of counters, I would add that I have seen examples of most of the problems described in this section in government count projects as well—including some in my own. Likewise, I hope that the generalizations on decision-making in the next section will be of value to government counters as well as to the others.

DISCUSSION AND CONCLUSIONS

In a free world, and particularly in the private sector, counters typically select methods that fit their particular objectives, and we all hope this freedom continues. Virtually all questions regarding methodology involve some qualification regarding availability of resources, as a universal method could only be applied given unlimited resources, if at all. Some counting exercises, e.g., aerial surveys of pelagic seabirds, may occupy 10 or more people, with annual budgets of hundreds of thousands of dollars; obviously this is not something for the spare-time amateur, and it was not possible at all until the environmental impact stakes became big enough to cover it. Any method that requires more than one person at a time, or calls for more specialized equipment than the now-ubiquitous binoculars, or for knowledge over and above bird identification, will be less generally usable than a simpler one:

> #1: the simpler methods suffer fewer constraints than more complex ones, always provided they are adequately standardized.

Rigorous standardization is possible but generally impracticable, and some compromise is essential. The "bottom line" has to be set in terms of comparability, and particularly of comparability between different observers. Results of any count conducted by amateurs for recreation are likely to be used mainly by persons other than the counters; and any long-term comparisons of data from permanent plots (or transects) are likely to involve counts by different observers. Most methods not involving measured areas (i.e., relative methods) are especially sensitive to observer variability, owing to individual differences in acuity of hearing or sight. More intensive methods, and particularly the mapping census method, tend to be more easily replicated by other people. As resources usually are insufficient to allow exclusive use of reproducible methods on extensive areas, it is decidedly preferable to use them to calibrate other less intensive methods, which can be used more widely, than to rely solely on the latter in order to sample larger areas. Thus,

> #2: selection of methods for standardization must include some that can be replicated by other people, even if these must be coupled with quicker but more wide-ranging methods in most cases.

The reliability of the calibration methods usually arises out of intimate knowledge and understanding of the area and its birds, rather than from confidence limits or statistical tests. But if resources allow coverage of replicated baseline plots, even the statisticians may be kept happy (cf. Owens and Myres 1973).

No perfect method exists, so new methods must be tried, and innovation goes on apace, often at the expense of comparability. The graduate student has to show that he (or she) has thought up something new, and the consultant trying to stay afloat in the economic maelstrom won't weep if his rivals cannot use his data. But even these realists need to compare their new results with those of others, so it is to their advantage not to exclude existing methodology. No amount of standardization will or should be allowed to hamper those hardy spirits who wish to experiment with new or modified methods, but they need to be discouraged from "throwing out the baby with the bath water." A number of workers (e.g., Bell et al. 1973, Best 1975) have suggested discarding one or another method because it did not deal effectively with some species or group in which they were interested. All existing methods suffer from some problems, but most have merit for some or most groups of birds. Where innovation is needed, even more than in developing wholly new methods (e.g., Emlen 1971, 1977a), is in *supplementing* existing methods for those "difficult" species or groups poorly sampled at present (cf. Erskine 1974). This conference restricted its discussion largely to "terrestrial" birds, presumably because many major groups of aquatic birds—seabirds, herons and other colonial water birds, waterfowl—are already subjects of voluminous census literature, with special conferences to discuss them. Birds that defend only the nest site, or rely on flight songs for advertising, or nest in colonies, give rise to many of the difficulties encountered with mapping censuses, as well as other intensive methods covering only small areas. Because each method was developed to take advantage of some aspect of bird behaviour that lends itself to systematic counting, we cannot be surprised if not all species share that particular aspect; birds have adapted to their environments by a wide diversity of behavioural patterns. Some methods are flexible enough to be used in many habitats and situations, alone or in combination with other methods, while others fit only a few, specialized species. Birds have shown themselves adaptable in exploiting diverse environments, and counters have to be adaptable in supplementing established methods where necessary to deal with a species poorly sampled by the method of one's choice. Therefore,

> #3: innovation in census methods should be encouraged, but especially to supplement existing methods rather than to replace them.

It should be preferable to retain the practicality and comparability of an established method side-by-side with an innovation to cover a particular case. For example, most people doing mapping censuses also count the nests of Starlings (*Sturnus vulgaris*) and swallows (Hirundinidae), since those species do not defend all-purpose territories.

Most of what I have said so far is just common sense, but it needs to be said once again lest it be forgotten amid the complications of modern science. We need clear, simple procedures, covering even quite elementary points, especially if we are to encourage participation by interested amateurs. The scientific method implies that if you describe precisely what was done it can be duplicated, but whoever drafted that creed had never heard of inter-observer variability. So we must also use methods that are not sensitive to changes in observers, in case someone else, sometime, somewhere, might want or need to use our data in comparisons. This of course is very charitable, good for the soul as well as for science, but how does it help achieve the objectives for which the census project was started? It reduces the options open to, and thus the decisions required of, the amateur who wants to feel that his (or her) hobby has some spin-off value. For other workers, its chief value lies in ensuring that comparative baseline data exist when they are needed. If everyone "does their own thing" in their own way, comparisons become nearly impossible, and no data base is accumulated. This is pretty well what did happen with many North American waterfowl surveys in the 1950s and early 1960s, since procedures were poorly standardized and often not even written down (cf. Diem and Lu 1960, Dzubin 1969). The accumulation of data banks depends on comparable results (e.g., Erskine 1980), which depend on standardized methods.

Finally, despite all my emphasis on comparability, I am not so naïve as to believe that standardizing methodology will always lead to comparable results. People working with census data also have to make decisions as to what can and cannot be compared. The mapping census gives density values with quite low inter-observer variability, but I showed earlier (Erskine 1974) that the numbers of breeding species claimed varied in a much more subjective fashion. Obviously, decisions to compare diversity indices (which are based on numbers of species as well as their density) for counts conducted by different observers are often on shaky ground. Compilers and analysts also need to ensure that the data they compare are comparable.