RELIABILITY OF SINGING BIRD SURVEYS: EFFECTS OF SONG PHENOLOGY DURING THE BREEDING SEASON

DAVID M. WILSON

AND

JONATHAN BART

ABSTRACT.—Singing bird surveys are frequently used to estimate the difference in density between two or more populations of a species detected almost entirely by vocalizations. Typically, the statistical analysis assumes that individuals in the populations have the same average probability of being detected. We estimated how much difference in average detectabilities might occur owing to variation in singing rates during the breeding season. Presence or absence of song proved to be the best measure of detectability. Among 125 House Wrens (*Troglodytes aedon*) in central Ohio, the probability of singing at least once during a 3-min period was 0.7 before mating, 0.5–0.6 from mating to the completion of egg-laying, 0.7 during incubation, and 0.5 or less thereafter. These probabilities were not detectably affected by season (1 June–15 July), time of day (first 5 h of daylight), or number of other nearby males (1 to 6). Our analysis shows that random phenological differences, such as an earlier season in one year as compared to another, might cause an error of up to 25% in the estimate of relative density. If several years of data are available, then the estimates of relative density will usually be more accurate. Substantial errors may still occur if changes in song phenology are density-dependent. Several methods have been designed to overcome the problem of unequal detectabilities, but most of them require the assumption that all birds close to the observer are detected; our study shows this assumption to be seriously in error for House Wrens. The study indicates that, in most cases, singing bird surveys should not be relied upon to produce accurate estimates of relative density.

All index methods require the assumption that individuals in the populations being compared have the same average probabilities of detection (Caughley 1977). For avian surveys in which birds are detected mainly by their songs, the equal-detectability assumption implies that the birds in each population must have the same average probability of singing at least once while the surveyor is present. Singing activity does not have to be the same for each bird; it may differ, for example, between mated and unmated birds. If such variation exists, however, then the frequency of birds in each song activity class must be about the same if the birds in the two populations are to be equally detectable.

To our knowledge, the only survey for which this issue has been investigated in detail is the Mourning Dove (*Zenaida macroura*) coo count, an annual survey coordinated by the United States Fish and Wildlife Survey. Baskett et al. (1978) reported that unmated doves appeared to sing far more actively than mated ones. Subsequently, Sayre et al. (1980) found that unmated males coo up to 20 times more often than mated males, and that the probability of cooing at least once during the 3-min survey period was nearly 1.0 for unmated birds,

but was only 0.15 for mated birds. If the proportion of mated birds at the time of the survey was 50% in one population and 75% in another population, then, applying the detectabilities above, the survey result from the second population would be 37% higher than that from the first population, owing solely to the difference in average detectability. Thus, for Mourning Doves, changes in reproductive status may cause substantial changes in survey results. This finding raises doubts about the effectiveness of these surveys in detecting changes in density.

It is well-known that singing activity varies during the reproductive season (i.e., Saunders 1951, Slagsvold 1977, Nolan 1978, Best 1981). Most such studies report a decline in song after mating, and a few of them provide quantitative estimates of this change. Von Haartman (1956), for example, found that unmated Pied Flycatchers (*Ficedula hypoleuca*) sing three times as many songs as mated ones, and Nice (1943) estimated that unmated Song Sparrows (*Melospiza melodia*) sing four times as often as mated ones. Best (1981) suggested that dramatic declines in song after pairing are characteristic of species for which the main function of song is mate attraction. He believed that song is more persistent when its function

TABLE 1. Variation during the nesting attempt of House Wrens in the number of songs per 3-minute period.

| | Stage of nesting attempt | | | | | | |
	ST	EN	LY	IN	NE	FL	UN
Mean	1.10	0.47	0.45	0.64	0.29	0.32	0.57
SE	(0.11)	(0.06)	(0.06)	(0.04)	(0.03)	(0.06)	(0.07)

ST = sticks (indicating unpaired male); EN = empty nest; LY = egg-laying; IN = incubation; NE = nestling period; FL = fledgling period; UN = after unsuccessful attempt.

is primarily territorial. These studies show that considerable variation may occur in the amount of song during the reproductive attempt, and that the pattern of variation probably differs widely among species. None of these studies, however, is sufficiently detailed to assess the possibility that surveys of singing birds may be monitoring reproductive success rather than, or in addition to, change in density.

We conducted a study to measure variation in singing activity during the nesting attempt and to determine how seriously such variation compromises the ability of singing bird surveys to measure changes in avian density. We selected the House Wren (*Troglodytes aedon*) for study because it is common and easily attracted to nest boxes, enabling us to measure and control density and to determine the reproductive status of many individuals.

STUDY AREA AND METHODS

The study was conducted on three areas in central Ohio farming country: Delaware Reservoir Wildlife Area, Alum Creek State Park, and private land surrounding the Village of Alexandria. We placed 102 nest boxes at 24 sites in each of the first two areas, and 204 boxes at 48 sites in the Alexandria area. The number of boxes per site varied from 1 to 6. At the multiple-box sites, boxes were placed approximately 60 m apart. Sites were at least 1.0 km apart to minimize interactions between male wrens at different sites. From 25 May until 15 July 1982, the nest boxes were checked at least once each week. Wrens were assigned to one of seven categories based on the contents of the nest box: sticks (indicating an unpaired male), empty nest, egg-laying, incubat-

ing, nestlings, fledglings, after unsuccessful nesting attempt. To measure song activity, we recorded the number of songs per bird during brief (6–12 min) periods during the first 5 h of daylight.

During the study, 240 of the boxes were used by 125 male House Wrens. Female wrens laid 97 clutches, 67 of which produced at least one fledgling. We monitored the male House Wrens for 5,808 min during 891 periods.

Data were analyzed using survey periods of 1, 3, 6, or 9 min. We were particularly interested in how variation in song activity would affect the Breeding Bird Survey (Bystrak 1981), so most of our analyses used a 3-min sampling period, as is used in that program. The significance level in all statistical tests was 0.01.

RESULTS

Males sang actively until acquiring a mate (Table 1). While the female lined the nest and began laying eggs, the male's song activity decreased steadily. During the egg-laying period, the average number of songs per 3 min was less than half that average before mating. During incubation, males sang significantly more often, although still substantially less than before mating. During the nestling period and after the young had left the nest, male song activity again diminished. The pattern of song activity during the nesting attempt was thus slightly bimodal, with a strong peak before mating and a weaker peak during the incubation period.

Detectability, defined as the proportion of periods with at least one song, showed this bimodal pattern more strongly (Table 2). With

TABLE 2. Variation during the nesting attempt of House Wrens in the probability that the male sings at least once during 1-, 3-, and 6-min periods.

| Interval length | | Stage of nesting attempt | | | | | | |
		ST	EN	LY	IN	NE	FL	UN
1-min	Mean	0.57	0.46	0.35	0.52	0.28	0.26	0.38
	SE	0.03	0.04	0.03	0.02	0.02	0.03	0.03
3-min	Mean	0.69	0.61	0.52	0.71	0.48	0.39	0.49
	SE	0.05	0.07	0.05	0.03	0.03	0.05	0.05
6-min	Mean	0.81	0.75	0.64	0.83	0.56	0.48	0.58
	SE	0.07	0.09	0.07	0.04	0.04	0.07	0.07

TABLE 3. Relationship between the probability that a House Wren sings at least once during a 3-min period (detectability) and time of day, time of season, and number of male conspecifics within 100 m.

Detect-ability	Time of day (hours after sunrise)				
	1	2	3	4	5
Mean	0.55	0.53	0.54	0.55	0.54
SE	0.05	0.05	0.05	0.06	0.16
	Number of conspecifics per site					
	1	2	3	4	5	6+
Mean	0.55	0.50	0.66	0.60	0.54	0.50
SE	0.02	0.03	0.03	0.06	0.07	0.08
	Time of year (month/day)					
	5/25–6/7	6/8–14	6/15–21	6/22–29	6/30–7/7	7/8–15
Mean	0.61	0.63	0.52	0.51	0.50	0.55
SE	0.05	0.03	0.03	0.04	0.05	0.04

1-, 3-, or 6-min listening periods, detectability was about equal prior to mating and during incubation and was substantially lower at other times. We investigated time of day and number of conspecifics at a site as possible confounding variables, but detectability did not show significant trends with either of these factors (Table 3).

DISCUSSION

To determine how much error could be caused by a difference in the average detectability of birds in two populations, it is easiest to imagine that the populations are of equal density. Several ways may then be used to estimate how much difference in survey results might be produced solely by differences in average detectability. For example, detectability (using a 3-min survey period) varied from approximately 0.7 (sticks, incubation) to 0.5 (after incubation), so it is mathematically possible for the survey results to differ by 40%. It is difficult to imagine this occurring in practice, because all the birds in one population would have to have completed incubation, and all the birds in the other population would have to be in the sticks or incubation stages. Nesting attempts are constantly being initiated during the season, so it would be rare for either population to be so highly synchronized.

A more realistic estimate of the possible error is provided by the seasonal change in detectability. During the first half of June, the proportion of birds singing at least once during a 3-min survey period was 0.62 (Table 3). In the second half of June, the proportion dropped nearly 25% to about 0.50. We conclude from this analysis that when two populations are being compared and one might be a few weeks ahead of the other phenologically, it is unsafe to infer a difference in density unless survey results differ by more than 25%.

When more than two populations are being compared, as occurs in long-term surveys or with several study areas along a habitat cline, then the situation is more complicated. If a change in average detectability is caused solely by chance factors, such as the season being earlier in one year than in another, then estimates of the long-term trend should not be affected seriously. Investigators using data from surveys such as the Mourning Dove coo count or the Breeding Bird Survey usually calculate trends on the basis of 10 or more years. It is difficult to see how chance factors could have much effect with so large a sample.

The changes in frequency of birds in each stage can be caused, however, not by random factors, but by the change in density, or by the same factor which causes the change in density. Assume, for example, that the survey results could be modelled using only two stages, pre-hatching and post-hatching (detectabilities 0.7 and 0.5, respectively), and that initially the population included 100 male House Wrens with 50 in each stage. The number of birds detected, sampling error aside, would be 60. Now suppose a predator or competitor invades the area causing mating or nesting failure before hatching, which results in (a) an increase in the proportion of birds in the pre-hatching stage from 50% to 60%, and (b) a 3% decline in the number of males present. In this hypothetical case, the number of birds detected in the survey after the decline would still be 60, and thus the decline would not be detected. If the adverse effect on this population continues, then the number of birds present will continue to decline and the proportion of them in the pre-hatching stage will continue to increase. Before many years pass, however, the survey will begin to reveal the decline. For example, when the number of males reaches 80, a decline will be evident even if all the males are in the pre-hatching stage (the number recorded would then be 56), and thereafter the survey will estimate the decline quite accurately. Two conclusions may thus be reached about factors which affect average detectability in a density-dependent manner: first, they may cause the survey to miss or underestimate a decline (or increase) for a few years; second, eventually the decline will be revealed, although its overall magnitude may be underestimated.

In the analysis so far, we have assumed that all birds which sing are detected. In fact, however, even the best surveyors fail to record many audible birds (Bystrak 1981, Bart and Schoultz 1984). If the assumption of recording all audible birds is relaxed, then it is not immediately obvious what measure of singing ac-

tivity is most appropriate for an evaluation of singing birds surveys. Mean number of songs is not the right measure, since the probability of detection certainly does not have a linear relationship with singing frequency. For example, actual detectability is probably about the same whether a bird sings eight times or four times per minute despite the 50% difference in number of songs. This is an important point since, other than Sayre et al. (1980), the few published estimates of stage-specific singing rates report only the number of songs per minute. The error, which can be caused by equating change in mean number of songs with change in detectability, can be appreciated by considering Sayre et al.'s (1980) data. Mean number of songs per 3 min changed 20-fold from before to after mating, whereas the probability of singing at least once changed only 7-fold.

Conversely, however, the probability that a given bird will be recorded is certainly higher if it sings 10 times than if it sings only once. There is thus a possibility that birds in two stages might have different actual detectabilities, despite having equal probabilities of singing at least once, owing to the mean number of songs per period being higher in one stage. Algebraically, the issue is:

$$\Sigma\, f_{1i}/\Sigma\, f_{2i} \quad \text{vs.} \quad \Sigma\, p_i f_{1i}/\Sigma\, p_i f_{2i},$$

where f_{1i} = P(that a bird sings i times in stage 1), i = 1, . . .
p_i = P(that a bird will be detected given that it sings i times).

The above ratio on the left is the estimated change in survey results using presence of song; the ratio on the right takes account of changes in detectability owing to change in the number of songs per period. The two expressions are nearly equal if either of two conditions is satisfied for all i: (1) f_{1i} is nearly equal to f_{2i}, or (2) p_i is close to the average detectability. We know of no reliable estimates of the detectability of birds as a function of the frequency with which they sing, but an examination of the data collected in our study shows that the frequencies of 1, 2, and 3 or more songs change little between stages, thus satisfying condition 1 above (Fig. 1). Probability of singing during the survey period thus appears to be the best available measure of change in detectability, and it is certainly better than change in mean number of songs.

ABILITY OF SURVEYORS TO ESTIMATE ABSOLUTE DENSITY

This study provides an opportunity to evaluate the validity of survey methods requiring detection of all individuals "close" to the sur-
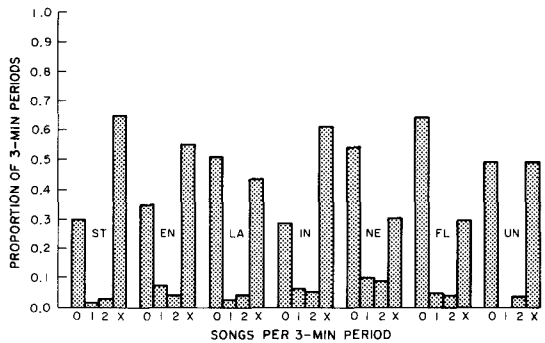

FIGURE 1. Variation during the nesting attempt of 3-min periods with 0, 1, 2, and 3+ songs. Abbreviations as in Table 1.

veyors. Three common methods make this assumption: Emlen's (1971) strip transect, Burnham et al.'s (1981) Fourier series method for line transect data, and Ramsey and Scott's (1981) variable circular plot method. In each case, an attempt is made to convert index data into estimates of true density. If the critical assumption of nearly 100% (or known) detectability close to the observer is not met, then, as the above authors acknowledged, these methods do not produce valid density estimates. Instead, they produce indices to density which may be no better than the original counts for monitoring population trends. They are certainly more complex to calculate and interpret. It is, thus, important to determine whether the assumption of high detectability is met for a variety of species.

In these methods, birds are counted while the surveyor either walks slowly through the area or stands at recording stations for a predetermined amount of time, usually between 3 and 9 min (see Ralph and Scott 1981). If we assume that surveyors who are walking can record any House Wren singing within 80 m (which we feel is a maximum distance), then a given bird could be within hearing distance for up to 6 min (assuming a walking speed of 1.6 km/h). Thus, a range of 1–9 min probably includes the entire time surveyors are within audible distance of each House Wren. It should also be noted that secretive species like House Wrens are not likely to be detected unless they sing.

Most of our recording periods were 6 min long. The probabilities that birds would sing at least once during 1-, 3-, or 6-min periods never exceeded 0.83 (Table 2). In a smaller sample ($n = 101$) of 9-min recording periods, the bird sang at least once in 81% of the periods. The probabilities of singing in the pre-incubation, incubation, nestling, and post-nestling stages were 0.87, 0.86, 0.82, and 0.76, respectively. These results indicate that the ab-

solute density of House Wren populations cannot be estimated reliably with any of the above methods unless the recording period is considerably longer than those commonly used.

FUNCTION OF SONG IN HOUSE WRENS

The findings that singing activity was not affected by the number of nearby conspecifics, and that it increased during incubation, support the view that mate attraction is a major function of song in House Wrens. Several authors have reported that singing rates of polygamous species (including House Wrens) do not decline after pairing, whereas songs of monogamous species do (Armstrong 1955, Andrew 1961, Verner 1963). At least three points should be made about these suggestions. First, the pattern we found—decrease after pairing, increase during incubation, decrease thereafter—is more complex than that envisaged by these authors. Second, although House Wrens are usually described as polygamous, only a few males were successful polygamists in this study, and Kendeigh (1941) reported only 6% polygamy in his long-term study of House Wrens. Third, the distinction between single- and double-brooded species may be as important as, or even more important than, the distinction between monogamy and polygamy. If song does serve primarily for attracting mates, at least after the start of the season, then perhaps song activity persists after pairing in both polygamous and double-brooded species and declines sharply after pairing in monogamous, single-brooded species.

ACKNOWLEDGMENTS

LITERATURE CITED

ANDREW, R. J. 1961. The displays given by passerines in courtship and reproductive fighting: a review. Part II. Ibis 103:549–579.

ARMSTRONG, E. A. 1955. The wren. Collins, London.

BART, J., AND J. D. SCHOULTZ. 1984. Reliability of singing bird surveys: changes in observer efficiency with avian density. Auk 101:307–318.

BASKETT, T. S., M. J. ARMBRUSTER, AND M. W. SAYRE. 1978. Biological perspectives for the Mourning Dove call-count survey. Trans. N. Am. Wildl. Nat. Resour. Conf. 43:163–180.

BEST, L. B. 1981. Seasonal changes in detection of individual bird species, p. 252–261. In C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. Avian Biol. 6.

BURNHAM, K. P., D. L. ANDERSON, AND J. L. LAAKE. 1981. Line transect estimation of bird population density using a Fourier series, p. 466–482. In C. J. Ralph and J. M. Scott [eds.]. Estimating the numbers of terrestrial birds. Stud. Avian Biol. 6.

BYSTRAK, D. 1981. The North American Breeding Bird Survey, p. 34–41. In C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. Avian Biol. 6.

CAUGHLEY, G. 1977. Analysis of vertebrate populations. John Wiley and Sons, New York.

EMLEN, J. T. 1971. Population densities of birds derived from transect counts. Auk 88:323–342.

KENDEIGH, S. C. 1941. Territorial and mating behavior of the House Wren. Ill. Biol. Monogr. 18:1–120.

NICE, M. M. 1943. Studies in the life history of the Song Sparrow. II. Trans. Linn. Soc. N.Y. 6.

NOLAN, V. 1978. The ecology and behavior of the Prairie Warbler, Dendroica discolor. Ornithol. Monogr. 26.

RALPH, C. J., AND J. M. SCOTT [EDS.]. 1981. Estimating the numbers of terrestrial birds. Stud. Avian Biol. 6.

RAMSEY, F. L., AND J. M. SCOTT. 1981. Analysis of the variable circular plot method, p. 483–487. In C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. Avian Biol. 6.

SAUNDERS, A. A. 1951. Bird Song. N.Y. State Mus. Handbook 7.

SAYRE M. W., T. S. BASKETT, AND K. C. SADLER. 1980. Radiotelemetry studies of the Mourning Dove in Missouri. MO Dep. Conserv. Terr. Ser. 9.

SLAGSVOLD, T. 1977. Bird song activity in relation to breeding cycle, spring weather, and environmental phenology. Ornis Scand. 8:197–222.

VERNER, J. 1963. Song rates and polygamy in the Long-billed Marsh Wren. Proc. 13th Int. Ornithol. Congr. (1962):299–307.

VON HAARTMAN, L. 1956. Territory in the Pied Flycatcher Muscicapa hypoleuca. Ibis 98:460–475.

Ohio Cooperative Wildlife Research Unit, Department of Zoology, The Ohio State University, Columbus, OH 43210.