The Condor 93:676–682

POINT COUNTS FROM CLUSTERED POPULATIONS: LESSONS FROM AN EXPERIMENT WITH HAWAIIAN CROWS[1]

GREGORY D. HAYWARD[2]
Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83848

CAMERON B. KEPLER
South East Research Group, U.S. Fish and Wildlife Service, School of Forest Resources, University of Georgia, Athens, GA 30602

J. MICHAEL SCOTT
Cooperative Fish and Wildlife Research Unit, U.S. Fish and Wildlife Service, Department of Fish and Wildlife Resources, University of Idaho, Moscow, ID 83843

Abstract. We designed an experiment to identify factors contributing most to error in counts of Hawaiian Crow or Alala (*Corvus hawaiiensis*) groups that are detected aurally. Seven observers failed to detect calling Alala on 197 of 361 3-min point counts on four transects extending from cages with captive Alala. A detection curve describing the relation between frequency of flock detection and distance typified the distribution expected in transect or point counts. Failure to detect calling Alala was affected most by distance, observer, and Alala calling frequency. The number of individual Alala calling was not important in detection rate. Estimates of the number of Alala calling (flock size) were biased and imprecise: average difference between number of Alala calling and number heard was 3.24 ($\pm$0.277). Distance, observer, number of Alala calling, and Alala calling frequency all contributed to errors in estimates of group size ($P < 0.0001$). Multiple regression suggested that number of Alala calling contributed most to errors. These results suggest that well-designed point counts may be used to estimate the number of Alala flocks but cast doubt on attempts to estimate flock size when individuals are counted aurally.

Key words: Point counts; census techniques; observer variation; flock size; Hawaiian Crow; Alala.

INTRODUCTION

Estimates of bird abundance are central to many ecological field investigations. Transect and point counts are frequently used to obtain abundance estimates for studies that explore trophic dynamics, habitat selection, and species interactions (Ralph and Scott 1981). Unfortunately an abyss often separates field counts and reliable estimates of bird abundance. In many applications of transect and point counts, the degree to which model assumptions are violated and the magnitude of sample bias are unknown (Verner 1981, 1985).

In the application of line transect and point counts to estimate population size we assume that no bird is counted more than once and that all birds on the line (or at plot center in point counts) are detected. Furthermore, detections must be independent events and populations that occur in groups must be counted as clusters (Burnham et al. 1980). To obtain an abundance estimate from clustered populations, cluster or flock size must be determined for each cluster detected.

When counts are used as an index to population abundance, fewer assumptions apply. In particular, estimates of flock size are not necessary if average flock size does not change over the inference period. Consistent under-counting or over-counting is not a problem if the ratio "number heard/number present" changes consistently with the number present. Unfortunately few indices have been validated, so many of the concerns related to reliable abundance estimates apply to unvalidated indices.

During field counts, observers detect birds and flocks of birds visually and aurally. In forested environments, most detections are for birds heard and not seen. Although an individual bird may vocalize more than once during a counting period, the observer must determine whether repeat vocalizations represent a new individual.

[1] Received 21 November 1990. Final acceptance 15 March 1991.

[2] Present address: Department of Zoology, Colorado State University, Fort Collins, CO 80523.

Scott and Ramsey (1981) noted the potential importance of "swamping" or saturation in underestimates of bird numbers. Systematic error from double counting or underestimating flock size could substantially bias samples and lead to erroneous estimates of population size. The potential magnitude of observer error in estimates of flock size has not received experimental study. In 1981, a working group charged with identification of research priorities in avian censuses listed studies of bias and variance in estimating numbers of birds as top priority (Verner 1981). Quinn (1981) explored the role of group size in determining detection probability. He noted that while detectability increases with group size, the relationship is not linear. Bart (1985) and Bart and Schoultz (1984) examined the effect of bird density on the accuracy of counts. Here we describe results of an experiment designed to determine the magnitude of observer error and to identify those factors contributing most to error in counts of Hawaiian Crow or Alala (*Corvus hawaiiensis*) groups that are detected aurally. Our analysis focuses on two broad questions; why do observers fail to detect some groups of Alala during a census, and when a group is detected, what are the characteristics of errors in estimates of group size.

STUDY SITE, SPECIES CHARACTERISTICS, AND METHODS

In 1978, the State of Hawaii maintained a small captive flock of Alala at the Endangered Species Breeding Facility at Pohakuloa, Hawaii, 2,100 m in the saddle between Mauna Kea and Mauna Loa. Three adults (two males, one female) and six juveniles (two 1-year olds, four young-of-the-year) were held in aviaries 6.1 m × 12.2 m × 3.7 m high. The chicken-wire pens were open to the sky but enclosed on their eastern sides by wooden panels. The facility was surrounded by an open to closed-canopy mamane (*Sophora chrysophylla*)-naio (*Myoporum sandwicense*) woodland (Scott et al. 1984). Topographically, the site is relatively flat to the south, east, and west, dropping about 43 m/km from east to west. The south slope of Mauna Kea begins its steep ascent from 2,100 m to 4,390 m (in 9.5 km) about 1 km north of the aviary. Winds, when present, consistently blew from the northeast.

The Alala, endemic to the island of Hawaii, is gregarious and vocal, similar to other Corvids (Munro 1960). The calls of adults and juveniles are similar—a harsh caw repeated rapidly. The Alala is the "noisiest bird in the lower Kona forests at daybreak" (Munro 1960) and the loud call can be heard from over 2 km. In the early morning, captive individuals called frequently.

To investigate the accuracy of counts of vocalizing Alala we conducted an experiment to simulate point counts. We established four 2,500-m transects extending from the Alala pens roughly in the four cardinal directions. Transects were marked with 25 stations at 100-m intervals. From 1–4 August 1978, several observers walked the transects. Observers (one to a station) listened for 3-min intervals at each station in sequence along a transect, recording all bird species detected and the number of Alala heard. Point counts extended from 05:20 to 08:00 hr on fair days with wind speed Beaufort 4 or less, to mimic bird censuses for ecological investigations in the region. During the point counts, two observers recorded Alala calls at the captive pens. Another field person coordinated the 3-min point counts with the counts of Alala at the pens by using radio communication.

During each 3-min count, observers at the Alala pens recorded the number of adults and number of young Alala that called, the number of individual vocalizations for both adult and immature birds, and rated the intensity of calls as either soft, loud, or very loud. Personnel conducting the point counts recorded wind speed (Beaufort Scale), wind direction, distance from Alala pens, time, and the number of Alala detected.

Observers used for point count simulations were all highly trained field personnel (Kepler and Scott 1981) who had censused birds in Hawaii for at least 10 weeks prior to the experiment. Each observer had his or her hearing tested and exhibited a maximum loss of 20 db from 1 to 8 KHz and had heard Alala on numerous occasions.

We examined three measures of error in the perceived number of Alala calling: the difference between number recorded at the pen and number heard at point counts (DIFF), the ratio of number heard and number that called (RATIO), and a value that we called the proportional error (absolute value of [1 − HEARD/CALLING]) (PROERROR). The proportional error indicates the percent difference between the two values regardless of which value is larger. None of these measures seems superior but each provides a different view of how point counts were related to the number of Alala actually calling.

In addition to determining the magnitude of error in point counts and extent of bias, we sought to determine the variables that influenced error rate. Potential variables included observer differences, distance of observer along transect, number of Alala calling, Alala calling frequency, calling intensity, wind speed, and transect conditions. We approached this problem largely as a model-building exercise; we attempted to build simple models by using the measured variables to predict errors in the point counts.

We used logistic regression to determine how the measured variables influenced whether observers failed to detect the group of calling Alala. For this analysis the response variable (heard/not heard) was binomial, and potential explanatory variables were a mixture of categorical (i.e., observer, transect, calling intensity) and continuous (i.e., distance, calling rate, etc.) variables. To evaluate the contribution of the measured variables to errors in the estimates of group size, we examined multiple regression models for continuous variables and analysis of variance for categorical variables (or transformed continuous variables).

To reduce the number of potential explanatory variables we combined four transects and five recorded wind speeds (Beaufort Scale) into a four-level composite variable. When winds blew, they consistently came from the northeast. Aside from the steeply rising slope on the north transect, the orientation of the wind with respect to the Alala cage was the most obvious difference among transects. Therefore we grouped transect/wind combinations into four categories as follows: west transect on calm (Beaufort Scale 0, 1) days; north, east, or south transects on calm days; west and south transects on windy days (Beaufort scale >1); north or east transects on windy days.

RESULTS

PROBABILITY OF DETECTING ALALA FLOCKS

Observers failed to detect calling Alala in 197 of 361 trials. Detection failure occurred under a wide range of circumstances; calling Alala were missed from all distances (100–2,500 m), by six of the seven observers, on all transects, across a range of Alala calling frequencies (8–169 calls in 3 min), and when the wind was both calm and blowing strongly. On an average, failures were 1,216 ($\pm$73.5) m from the Alala cage, when five ($\pm$0.21) birds were calling, and 62 ($\pm$5.1) individual calls were given during the 3-min count period. The wind was not blowing on 77% of the occasions.

Although most failures occurred at distant listening stations, in 12 trials observers failed to hear Alala at 300 m or less (six at 300 m, five at 200 m, and one at 100 m). The wind was calm on all but one of these trials and calling intensity was rated loud in eight instances. In all 12 of the short-distance failures, three to seven Alala called and calling frequencies ranged from 20–81 ($\bar{x}$ = 46.4 $\pm$ 10.93) caws in 3 min. Five of seven observers used in the experiment missed groups of Alala calling from 300 m or less.

Line-transect and point-count methods assume that the probability of detection decreases as the distance to an object increases. The form of the detection function is important in determining the applicability of line-transect models to sample data (Burnham et al. 1980, Scott et al. 1986). In particular the detection rate is assumed to be a monotonically decreasing function of distance. We plotted Alala detection frequencies at increasing distances to assess the form of the detection function in these trials (Fig. 1, all trials). Aside from sample variation, the detection curve exhibited desired properties for point-count estimates of abundance; detection decreased with distance.

We examined several logistic regression models to determine the combination of variables, in addition to distance, that were important in the detection of a group of Alala. The models differed in the number of levels of several categorical variables and in the number of variables included in the models. All models lead to similar conclusions, but it was useful to explore several models to examine trade-offs between increased statistical power (with few variables) and a broad view (with many variables). A model including all measured variables indicated that distance from cage (P = 0.0001), observer (P = 0.0003), and Alala calling rate (P = 0.022) were most important, while calling intensity (P = 0.062), transect/wind interaction (P = 0.873), and the number of Alala calling (P = 0.941) were less important.

The insignificant role of Alala group size (number calling) in determining group detection was surprising. A model including only distance and number of Alala calling emphasized the subor-

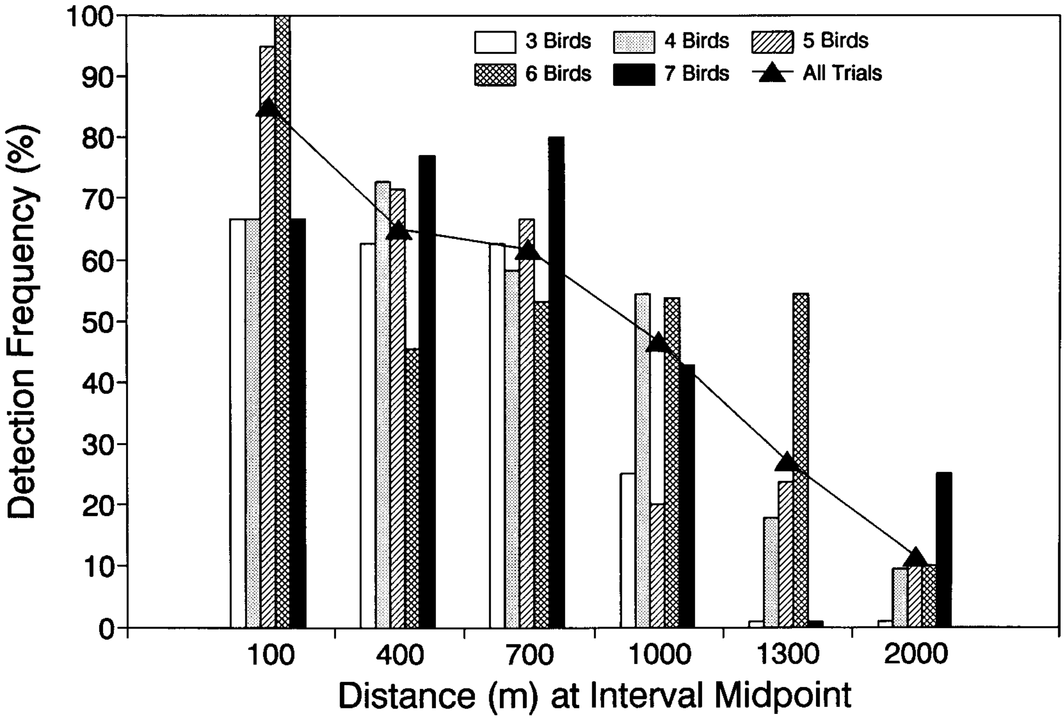

FIGURE 1. Sample detection curve for a flock of Alala detected aurally in 3-min point counts from stations 100–2,500 m from a captive flock of Alala. Data for flocks of three to seven Alala are displayed separately as bars along with pooled results.

dinate role of group size. Chi-square statistics were 69.1 (df = 1) for distance and 12.93 (df = 7) for number calling, suggesting that distance is more important in predicting whether a group will be detected.

The role of group size in determining detection frequencies is illustrated in Figure 1. In spite of substantial scatter due to sample variation, detection curves for different group sizes have similar shapes, indicating no tendency for larger groups to be more easily detected aurally. In fact, the slopes of linear regressions relating detection rate to group size for six different distance categories (0–200 m, 300–500 m, 600–800 m, 900–1,100 m, 1,200–1,400 m, ≥1,500 m) were significantly different from zero in just one situation ($P = 0.61$; 0.97, 0.42, 0.57, 0.67, 0.04 for the five distance categories respectively (see Fig. 1 for pattern)).

PRECISION AND BIAS IN ESTIMATES OF GROUP SIZE

After determining what variables influenced detection of Alala groups, we examined the success of observers estimating the number of Alala calling in those groups that were detected. We studied three measures of estimate error. DIFF indicates the magnitude and direction of errors in number of birds counted. If counts are not biased, we expect the average DIFF to equal zero. Similarly, the average RATIO will equal "1" if estimates are not biased. Bias cannot be detected with PROERROR, which measures accuracy. The variance or coefficient of variation of each of these measures may be used as an indication of precision.

We explored the magnitude of error and extent of bias using only those observations when at least one crow was detected (Table 1). During line-transect counts, groups of birds that are not on the line or directly overhead can be undetected without biasing estimates of abundance. In other words, the method assumes that some groups will not be detected and the probability of detection decreases with distance. When a group is detected, however, group size must be estimated without bias and with high precision to obtain reliable estimates of species abundance.

TABLE 1. Three measures of error in aural estimates of group size for Alala detected from listening stations from 100–2,500 m from a caged breeding flock. Group size is defined as the number of Alala that vocalized during a 3-min sample period. See METHODS for descriptions of error metrics.

Error measure	Distance (m) 0–500	600–1,000	1,100–1,500	>1,500
DIFF (mean)	2.49	3.59	4.58	4.14
(SE)	0.190	0.216	0.284	0.508
RATIO	0.55	0.40	0.23	0.29
	0.033	0.029	0.019	0.043
PROERROR	0.44	0.61	0.77	0.71
	0.026	0.025	0.019	0.043

Observers both underestimated and overestimated the number of Alala calling during 3-min observation periods but overestimates occurred in only three situations. In 157 observations from 1,500 m or less, the average difference between the number of crows actually calling and the number heard was 3.24 (±0.277); coefficient of variation (CV) was 54.7%. For this same sample, RATIO averaged 0.44 (±0.041) with a CV of 59%, while PROERROR averaged 0.58 (±0.034) with a CV of 39%. These results suggest a strong negative bias; observers consistently underestimated the number of crows calling (Table 2). Precision was also low, as indicated by PROERROR, which suggests that estimated group size averaged only within 60% of the true value.

What variables contributed to errors in estimates of group size? It will not surprise most students of bird census techniques that most measured variables contributed to estimate errors. Results of an ANOVA cast as a fixed-effects, factorial design with six treatments suggested that distance, observer, number of Alala calling, and

TABLE 2. Relation between the number of Alala perceived by field observers during 3-min sample periods and the number of captive Alala observed calling.

Number heard	Number of Alala calling 2	3	4	5	6	7	8	9
0	5	20	47	45	42	29	4	5
1	0	5	12	13	10	3	5	0
2	0	3	11	12	11	16	7	0
3	0	4	4	7	4	12	0	0
4	0	2	1	7	6	2	0	1
5	0	1	0	1	2	2	0	0
Total	5	35	75	85	75	64	16	6

TABLE 3. Analysis of variance exploring the role of six factors in determining errors in estimates of Alala group size. RATIO, a measure of error (number heard/number calling), is the response variable and the analysis is cast as a fixed-effects, factorial design without interactions. Sample includes 128 observations when at least one Alala was heard.

Source	df	F	$P > F$	Differences[1]
Distance	5	6.22	0.0001	8
Observer	4	6.30	0.0001	3
Calling freq.	5	7.23	0.0001	10
No. calling	5	6.38	0.0001	10
Calling intensity	2	1.55	0.227	0
Transect-wind interaction	2	0.83	0.438	1

[1] Number of significant pairwise differences, Duncan's test.

calling frequency all contributed to error (Table 3). Calling intensity and the interaction of transect/wind were the only factors that could not be shown to significantly influence estimate error. Analysis of variance using the three measures of error (DIFF, RATIO, and PROERROR) gave qualitatively identical results.

Stepwise multiple regression examining all noncategorical variables (distance, frequency of calls, number of Alala calling, and wind speed) indicated that DIFF was most strongly related to the number of Alala calling (or group size). For this analysis we included squared terms for each of the four variables to account for nonlinear responses. Distance^2 entered the model second and calling frequency entered third. Partial R^2 values for the model with three terms were 0.71 for number of birds calling, 0.05 for distance^2, and 0.02 for calling frequency. These three variables explain more than 75% of the variation in the observed error as measured by DIFF.

We further examined the interaction between number of Alala calling (group size) and error in estimates of group size. Errors grew as the size of the group being observed increased. The average difference between observed and actual group size rose from 0.6 to 6.4 as the number calling changed from three to eight. A simple linear regression of DIFF × CALLING showed a slope near 1 ($b = 0.97$, $P = 0.0001$) and regression using the other measures of error was significant at the same level. We examined plots of residuals from these analyses and saw no evidence that nonlinear models would describe the relation more clearly.

The variance in error did not increase with group size. In fact, the coefficient of variation for all three measures of error declined as group size increased from three to eight.

DISCUSSION

The task of estimating abundance of flocking birds such as Alala involves estimating the number of flocks and the distribution of flock sizes. During both steps numerous variables may influence the accuracy of sample estimates. Some of the confounding variables can be controlled through rigorous sample design (i.e., diurnal variation) and personnel training, while others cannot (i.e., bird behavior, which influences detectability).

In this paper we have dealt specifically with estimates of population abundance based on aural detections. The sample frame therefore only includes those birds which call, while the parameter of interest is total population size. Evaluation of the estimator, then, must focus on the relation between the sampled population and the population of interest.

Our results implied that accurate estimates of the number of Alala flocks can be obtained from point-counts of calling birds. The form of the detection curve reflected the shape expected for estimates using line transect technology (Burnham et al. 1980). More important, the probability of detecting a group at a given distance was, surprisingly, not an increasing function of group size. When animals are *seen* in groups, the probability of sighting them is usually an increasing function of group size (Samuel et al. 1987) and modification of basic line transect formula is necessary to estimate the abundance of groups (Quinn 1981). The expected relation did not hold for Alala detected only through calls possibly because of differences in how humans perceive objects by sight and by hearing.

Three variables were most important in determining whether flocks were detected or not, distance, Alala calling frequency, and observer. Dealing with observer variation may be the most important task in controlling the quality of estimates of flock abundance because this variable can be influenced by experimental design. Verner and Milne (1989:198) suggest several ways to control observer variability. They stress using multiple, well trained, and tested observers who sample all stations. Results from the observers are pooled for analysis.

Although we found that estimates of the number of Alala flocks could be accurately estimated, determining group size using point counts of calling birds appears more difficult. Estimates of group size had a strong negative bias and were imprecise. Because of the problems associated with estimating group size using standard point count procedures, alternative approaches must be used if estimates of abundance, rather than an index, is needed. It may be necessary to estimate average group size using a separate survey in which flock size is recorded for flocks detected near the observer. In short, the sample frame must be changed so the population sampled fits more closely the population of interest.

What do these results indicate about previous estimates of the endangered Alala population (Giffin et al. 1987, Scott et al. 1986)? Because of the negative bias in estimates of group size, previous estimates were likely underestimates of population size even though counts occurred when Alala were least likely to occur in flocks. The magnitude of the bias is unknown because flock size was estimated using a combination of aural and ocular detections. Similarly, estimates of wintering passerine populations and flocking game birds, which rely on aural detection, should be critically reviewed in light of these results.

ACKNOWLEDGMENTS

This work was made possible through use of captive Alala at the Endangered Species Breeding Facility at Pohakuloa, Hawaii and through assistance from staff at the facility and the Hawaiian Department of Natural Resources. We thank the bird survey crew including P. Ashman, C. Atkinson, T. Burr, T. Casey, H. Hunt, R. McArthur, and P. Pyle. Reviews by J. Bart and J. Verner were especially helpful. This contribution is no. 578 of the Idaho Forest Wildlife and Range Experiment Station, and was funded in part by the Idaho Cooperative Fish and Wildlife Research Unit and Idaho Department of Fish and Game.

LITERATURE CITED

BART, J. 1985. Causes of recording errors in singing bird surveys. Wilson Bull. 97:161–172.

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

BURNHAM, K. P., D. R. ANDERSON, AND J. L. LAAKE. 1980. Estimation of density from line transect sampling of biological populations. Wildl. Monogr. 72:1–202.

GRIFFIN, J. G., J. M. SCOTT, AND S. MOUNTAINSPRING. 1987. Habitat selection and management of the Hawaiian crow. J. Wildl. Manage. 51:485–494.

KEPLER, C. B., AND J. M. SCOTT. 1981. Reducing count variability by training observers, p. 366–371. *In* C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. Avian Biol. No. 6.

MUNRO, G. C. 1960. Birds of Hawaii. Charles Tuttle Co., Rutland, VT.

QUINN, T. J. 1981. The effect of group size on line transect estimators of abundance, p. 502–508. *In* C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. Avian Biol. No. 6.

RALPH, C. J., AND J. M. SCOTT [EDS.]. 1981. Estimating numbers of terrestrial birds. Stud. Avian Biol. 6:1–630.

SAMUEL, M. D., E. O. GARTON, M. W. SCHLEGEL, AND R. G. CARSON. 1987. Visibility bias during aerial surveys of elk in northcentral Idaho. J. Wildl. Manage. 51:622–630.

SCOTT, J. M., S. MOUNTAINSPRING, F. L. RAMSEY, AND C. B. KEPLER. 1986. Forest bird communities of the Hawaiian Islands: their dynamics, ecology, and conservation. Stud. Avian Biol. 9:1–429.

SCOTT, J. M., AND F. L. RAMSEY. 1981. Effects of abundant species on the ability of observers to make accurate counts of birds. Auk 98:610–613.

SCOTT, J. M., S. MOUNTAINSPRING, C. VAN RIPER III, C. B. KEPLER, J. D. JACOBI, T. A. BERR, AND J. G. GRIFFIN. 1984. Annual variation in the distribution, abundance, and habitat of the Palila (*Loxioides baileui*). Auk 101:647–664.

VERNER, J. 1981. Appendix VI: Report of the working group to identify future research needs, p. 548. *In* C. J. Ralph and J. M. Scott [eds.], Estimating the numbers of terrestrial birds. Stud. in Avian Biol. No. 6.

VERNER, J. 1985. Assessment of counting techniques, p. 247–302. *In* R. F. Johnston. [ed.] Current ornithology, Vol. 2. Plenum Press, New York, N.Y.

VERNER, J., AND K. A. MILNE. 1989. Coping with sources of variability when monitoring population trends. Ann. Zool. Fennici. 26:191–199.