SAMPLING BIAS IN DISPERSAL STUDIES BASED ON FINITE AREA

By George F. Barrowclough

INTRODUCTION

Dispersal is the movement made by an individual animal from its birthsite to the place at which it reproduces (Howard, 1960). Dispersal, so defined, is directly related to gene flow (Crumpacker and Williams, 1973; Powell et al., 1976; Endler, 1977). Since gene flow may be of major importance in evolutionary processes (Mayr, 1963; Slatkin, 1973; Endler, 1977), analysis of dispersal is of considerable importance to students of evolutionary biology.

In birds, dispersal is most conveniently measured through the recovery, at their nesting sites, of birds originally banded at their birthsites. If a large number of such returns are obtained, a distribution can be prepared of the number of returns as a function of dispersal distance. It is from this distribution that parameters pertinent to gene flow are obtained. For example, variance and kurtosis, both important concepts in population genetics (Wright, 1969), can only be extracted from the distribution of dispersal distances. Mean recovery distance is not adequate.

Since it is recoveries of birds at their nesting sites, and not all recoveries, that are important for evolutionary genetic purposes, the best source of avian dispersal information appears to be those studies aimed at banding and recovering all individuals of a species within an area over a period of years. Examples of this are Nice's (1937) work on the Song Sparrow (*Melospiza melodia*), Kendeigh's (1941) study of the House Wren (*Troglodytes aedon*), and Kluijver's (1951) paper on the Great Tit (*Parus major*).

The purpose of this paper is to develop a correction for a systematic bias inherent in dispersal distributions obtained from some such intensive local banding and recapture studies. The problem arises because the probability of observing a dispersal event of distance x is inversely related to that distance. This is obvious for large values of x. For example, if x is greater than the longest dimension of the study site, then the probability of observing such an event is zero, unless recoveries by the general public and other banders are used. Such recoveries will be biased, however, unless the potential for these recoveries is uniformly distributed with respect to angle and distance from the study site, and is as high outside the study area as it is inside. Since this generally will not be the case, only recoveries within the study area should be used.

For distance x less than the dimensions of the study site, it has not been generally recognized that a bias still exists. This is shown in Figure 1. In a study area of arbitrary shape, as shown, the probability of detecting short dispersal events is nearly one for birthsites, such as A, in the center of the study area. For birthsites closer to the edge of the
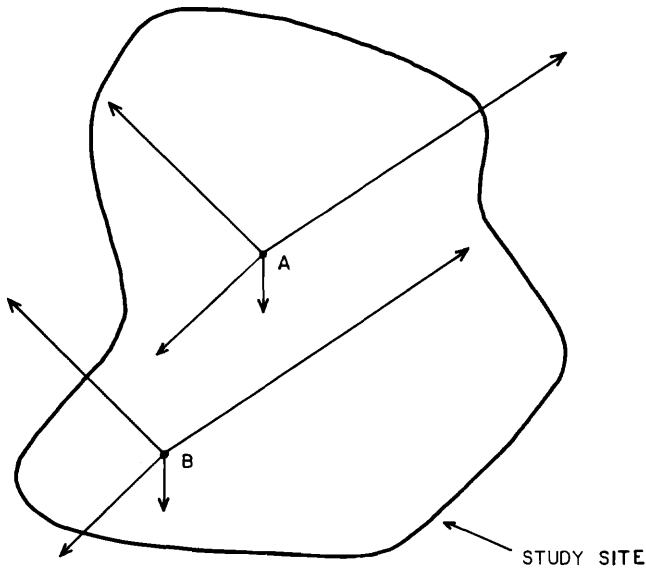
333

FIGURE 1.   For a finite study area, the probability of observing a dispersal event of distance
x (indicated by arrows) depends upon the location of the birthsite (A vs. B) within
the study area.

study area, such as B, the probability of observing short dispersal distances is less than one. In general, the probability of detecting dispersal distances of length x decreases as the site of origin is moved towards the boundary of the studied area, for dispersal distances less than the approximate radius of the study area. For distances larger than this, dispersal events can only be observed for sites of origin away from the center of the area. In the limit, dispersal distances of approximately the same length as the study site could only be observed if they happened to be from near one boundary of the site to a point near the opposite boundary. Longer dispersals are never observed.

Thus the probability of detecting dispersal events is a function of the distance dispersed and the site of origin within the study area. Consequently the distribution of dispersal distances, if uncorrected for this sampling bias, will reflect a relative excess of short distance dispersals and a deficit of longer ones.

METHODS

The systematic bias discussed above can be removed if the observation of dispersal distances is treated as a problem in sampling an unknown distribution. If we obtain the probability of observing a dispersal event of distance x, as a function of that distance, as well as of the study area and the site of origin within the study area, then the actual number of

dispersers moving a distance x can be estimated as the observed number divided by the probability of observing such events. For example, if eight birds are observed to have dispersed a distance of 500 to 600 m, and the probability of observing dispersal events of this distance interval is 0.5, then our estimate of the true number of dispersers in this distance class is 16.

The probability of observing a dispersal event of distance x can be computed for a simplified model as follows: it is assumed that a study site can be reasonably approximated as a circle of radius R. If birds raised at a nest site a distance r from the center of the study area disperse a distance x, with no directional bias, Figure 2a, then a fraction $(\pi - \theta)/\pi$ of such dispersers will remain within the study area, and hence be observed. To obtain θ in terms of R, r and x, we first note that $\theta = \cos^{-1}(y/x)$, Figure 2b. In addition $x^2 = z^2 + y^2$ and $R^2 = z^2 + (r + y)^2$. Consequently, $y = (R^2 - x^2 - r^2)/2r$. Thus, the probability of observing a dispersal event, given R, r and x, can be written $\{\pi - \cos^{-1}[(R^2 - x^2 - r^2)/2xr]\}/\pi$. Call this quantity $P(R,x,r)$. Note that r can vary from $R - x$ to R. For r between zero and $R - x$, all dispersers are observed in exhaustive studies; hence in that region let $P(R,x,r) = 1$. If it is assumed that nest sites are uniformly distributed over the area of study, the proportion of observed dispersers moving a distance x is equal to the integral of $P(R,x,r)$ over the area of the study site, normalized by the total area of the study site. That is,

$$P'(R,x) = \int_0^R \int_0^{2\pi} P(R,x,r) \, r \, d\theta \, dr/\pi R^2.$$

In practice the number of dispersers recovered in a study is always a small number. Only in the largest studies does it ever exceed 100 individuals. Consequently, the results of dispersal studies are often summarized by the number of dispersers moving a class of distances, i.e. 0–99 meters, 100–200 m, etc. In order to correct data summarized in this manner for finite area bias we need only integrate $P'(R,x)$ over the appropriate class of distances and normalize. Thus, the probability of observing dispersal events in the distance interval x_1 to x_2 is: $P''(R,x_1,x_2) = \int_{x_2}^{x_1} P'(R,x) \, dx / \int_{x_2}^{x_1} dx$. Finally the true number of dispersal events in this interval is estimated as the observed number divided by $P''(R,x_1,x_2)$.

The form of the definite integral $P''(R,x_1,x_2)$ could not be located in tables; in addition it was found that the expansion of the arc cosine in terms of its argument had poor convergence properties. Therefore a computer program has been written to evaluate the above probability by double numerical integration. For convenience the integration over the dispersal distances was broken up into two intervals, $0 \leq x \leq R$ and $R < x \leq 2R$. The exact form of the integral as evaluated is:
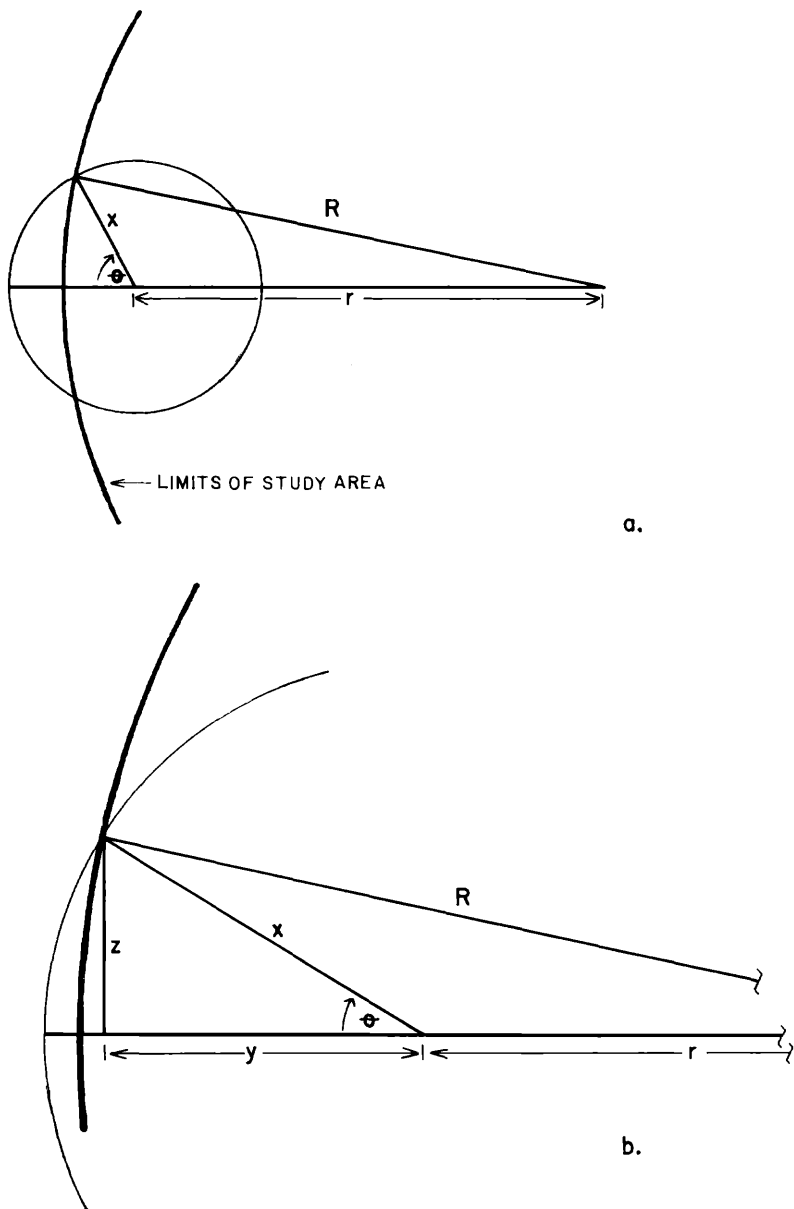
FIGURE 2.   Given a study site of radius R, if dispersers originating a distance r from the center disperse a distance x, then (a) a fraction $(\pi - \theta)/\pi$ of them will leave the study area. θ can be expressed in terms of R, r and x, by analyzing the x, y, z triangle (b), see text.

$$1.0 - 2/[\pi R^2(x_i - x_j)] \int_{x_j}^{x_i} \int_{R-x}^{R} \cos^{-1}\{(R^2 - x^2 - r^2)/2xr\}r \, dr \, dx$$

$$\text{(for } x \leqslant R\text{),}$$

$$+ \, 2/[\pi R^2(x_i - x_j)] \int_{x_j}^{x_i} \int_{x-R}^{R} \cos^{-1}\{(r^2 + x^2 - R^2)/2xr\}r \, dr \, dx$$

$$\text{(for } R < x \leqslant 2R\text{).}$$

A copy of the Fortran program used to correct recovery data for sampling bias is available from the author, but some changes will be necessary to adapt the program to numerical integration routines available at other institutions.

RESULTS

The corrected dispersal distributions have been computed for the dispersal data for House Wrens reported by Kendeigh (1941). The study area was treated as a circle having an area identical to the area of the Outfield, Ohio study site (Kendeigh, 1941, Fig. 4). The original and the corrected distributions are illustrated in Figure 3 for nestlings and for adult males. The distribution for adult females is intermediate between these two. The parameter of interest for studies of evolutionary genetics, the root-mean-square dispersal length (Crumpacker and Williams, 1973), is increased in each case. For the nestlings the correction increases the length by 23.2%. For the adult males, the uncorrected data underestimate the same parameter by 21.4%. It may seem surprising that the apparently small correction in the case of the adult males results in a change as great as 20%. This occurs because the calculation of the root-mean-square length gives much more weight to dispersers going long distances than it does to short distance dispersers. For the case of the male wrens, the last two distance classes, after correction, contain approximately twice as many individuals as they did before the correction. It is this increase that is responsible for the 20% change.

DISCUSSION

The problem addressed here has been recognized for some time. Kluijver (1951) and Bulmer (1973) were aware of it, as apparently was Wolfenbarger (1946). It was not addressed, however, in the early, classic studies of dispersal (Dobzhansky and Wright, 1943, 1947), because in those studies marked *Drosophila* were all released from a common central point. Thus there was no need for a correction for multiple sites of disperser origin within the study area. Recently, however, an increased interest in gene flow and genetic population structure in birds has occurred (Baker, 1975; Handford and Nottebohm, 1976). Consequently a knowledge of dispersal distributions for several avian species is desirable for comparison with such evolutionarily interesting cases as birds with dialects, communal nesters, and species with helpers. It is hoped that the correction developed here will facilitate such analyses. It also
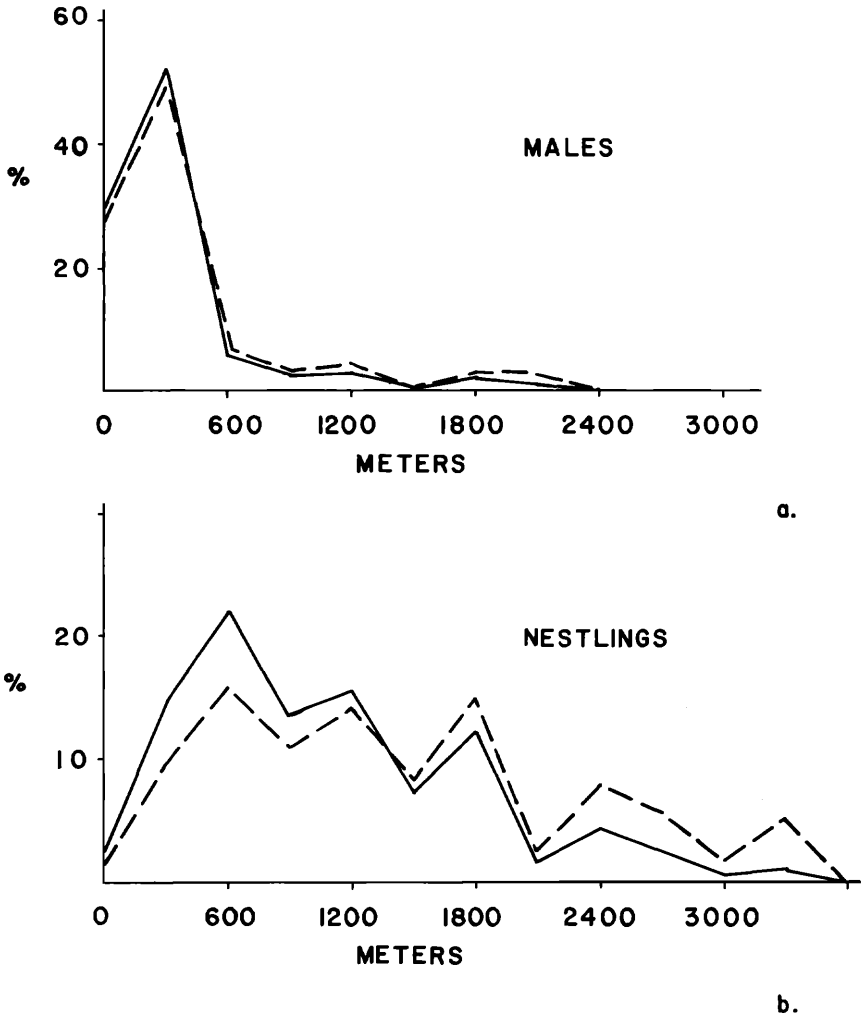
FIGURE 3. Uncorrected (solid lines) and corrected (broken lines) dispersal distributions for male (a) and nestling (b) House Wrens. Data from Kendeigh (1941).

should be noted that although genetics has been emphasized throughout this paper, the application of the correction is not limited to this topic. The bias is present in all studies of dispersal that involve finite study areas. Consequently the method may be of value to research in population dynamics, demography, and other fields in which individual movements play an important role.

It has been assumed that most study sites can be reasonably approximated by a circle with an area identical to that of the actual study site.

A method exists to check the adequacy of this approximation. In addition to the circle having identical area to the actual site, it is possible to compute a corrected dispersal distribution for the largest circle that can be inscribed within the study site and for the smallest circle that will circumscribe the site. The former circle generates a corrected distribution larger than the true one since some long distance dispersers actually are observed that this approximation assumes are missed. Hence this represents an upper bound on the true distribution. Likewise, the circumscribed circle produces a lower bound on the true distribution. If these bounds vary greatly from the equal area correction, then the possibility is suggested that the study site may not be represented well by a single circle. However, a correction can be applied to such grossly noncircular study sites provided the area is represented as a series of circular subareas. In this manner corrected distributions can be obtained for arbitrarily shaped sites. For such cases, however, the integration must be extended across subareas. The angular dependence, suppressed in the original development, has to be retained for such cases.

The correction factor for multiple circular areas is found by a method very similar to the geometrical analysis already described. The proper correction is the average of the corrections for the individual subareas, weighted by their relative areas, plus a between-circle term. This term involves the solution of the triple integral:

$$2/[R_1^2(x_1 - x_2)] \int_{x_2}^{x_1} \int_0^{R_1} \int_0^{\pi} \cos^{-1}\{(x^2 + [z^2 + r^2 - 2rz \cos \phi] - R_2^2)/$$
$$(2x\sqrt{[z^2 + r^2 - 2rz \cos \phi]})\} \, r \, d\phi \, dr \, dx,$$

where R_1 is the radius of the circular subarea the dispersers leave, R_2 is the radius of the subarea at which the dispersers arrive, z is the distance between the centers of the subareas, and r and ϕ are the polar coordinates of the site of origin of the dispersers in the first subarea. This integral must be computed in both directions for all pairs of subareas. The average of these between subarea terms, weighted by the relative sizes of the subareas, is added to the within subarea correction to give an overall correction factor. A computer program to compute the above integral has been prepared by the author.

A method for correcting dispersal data, as described here, will only be of interest if, in practice, it results in a significant modification of the dispersal distribution. Since the area involved in Kendeigh's study was quite large, the effect of the correction was not as great as it would be for a smaller study site. Nevertheless, that the correction was as large as 20% suggests that the effect of the modification generally will not be trivial. It should also be noted that if the numbers of individuals in the longer distance dispersal classes are small, so that some distance classes contain no individuals and adjacent classes contain only one or two, then any correction will result in large interclass differences. For such cases,

it is recommended that classes be pooled to contain 5+ individuals in order to reduce sampling error.

After the corrections discussed in this paper are made, it is possible that the proportions of dispersers in the longest distance classes that can be measured in a given study will not be approaching zero. This suggests that a significant part of the true dispersal distribution was not observed. If this is the case then it further suggests that the study area was not sufficiently large to yield a good estimate of dispersal parameters. Hence, the method outlined here can be used to check the validity of the estimated dispersal distribution. This cannot be done with an uncorrected dispersal distribution. In such cases the number of observed dispersers may approach zero, even if the study area is considerably smaller than some portion of actual dispersal distances, merely because the probability of observing longer distances falls to zero due to finite sampling area.

SUMMARY

In dispersal studies that involve banding and recovery of birds over a finite study area, a bias is found in the distribution of dispersal distances due to a non-uniform probability of recovery as a function of distance dispersed. A method is developed here to correct for this bias by computing the recovery probability as a function of dispersal distance, given information about the size and geometry of the study site. When applied to the House Wren data of Kendeigh (1941), the correction results in a 20% increase in the estimates of gene flow for that species. Although the method cannot correct for dispersal distances greater than the longest dimension of the study site, it can be used to evaluate better whether or not the study area was sufficiently large to yield a good estimate of dispersal parameters.

LITERATURE CITED

BAKER, M. C. 1975. Song dialects and genetic differences in White-crowned Sparrows (*Zonotrichia leucophrys*). *Evolution,* **29:** 226–241.

BULMER, M. G. 1973. Inbreeding in the Great Tit. *Heredity,* **30:** 313–325.

CRUMPACKER, D. W., AND J. S. WILLIAMS. 1973. Density, dispersion, and population structure in *Drosophila pseudoobscura. Ecol. Monogr.,* **43:** 499–538.

DOBZHANSKY, T., AND S. WRIGHT. 1943. Genetics of natural populations. X. Dispersion rates in *Drosophila pseudoobscura. Genetics,* **28:** 304–340.

———. 1947. Genetics of natural populations. XV. Rate of diffusion of a mutant gene through a population of *Drosophila pseudoobscura. Genetics,* **32:** 303–324.

ENDLER, J. A. 1977. Geographic Variation, Speciation, and Clines. Princeton, N. J., Princeton Univ. Press.

HANDFORD, P., AND F. NOTTEBOHM. 1976. Allozymic and morphological variation in population samples of Rufous-collared Sparrow, *Zonotrichia capensis*, in relation to vocal dialects. *Evolution,* **30:** 802–817.

HOWARD, W. E. 1960. Innate and environmental dispersal of individual vertebrates. *Amer. Midl. Nat.,* **63:** 152–161.

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

KLUIJVER, H. N. 1951. The population ecology of the Great Tit, *Parus m. major* L. *Ardea,* **39:** 1–135.

MAYR, E. 1963. Animal Species and Evolution. Cambridge, Mass., Harvard Univ. Press.

NICE, M. M. 1937. Studies in the life history of the Song Sparrow. I. A population study of the Song Sparrow. *Trans. Linn. Soc. N. Y.,* **4:** 1–247.

POWELL, J. R., T. DOBZHANSKY, J. E. HOOK, AND H. E. WISTRAND. 1976. Genetics of natural populations. XLIII. Further studies on rates of dispersal of *Drosophila pseudoobscura* and its relatives. *Genetics,* **82:** 493–506.

SLATKIN, M. 1973. Gene flow and selection in a cline. *Genetics,* **75:** 733–756.

WOLFENBARGER, D. O. 1946. Dispersal of small organisms. *Amer. Midl. Nat.,* **35:** 1–135.

WRIGHT, S. 1969. Evolution and the Genetics of Populations. Vol. 2. Chicago, Ill., Univ. of Chicago Press.