Studies in Avian Biology No. 6:483–487, 1981.

ANALYSIS OF BIRD SURVEY DATA USING A MODIFICATION OF EMLEN'S METHOD

FRED L. RAMSEY[1] AND J. MICHAEL SCOTT[2]

ABSTRACT.—This paper describes in general terms the data analysis procedures followed for the Hawaiian Forest Bird Survey. The method consists of first examining detection distances to estimate the Effective Areas Surveyed—a modification of Emlen's Coefficient of Detectability, then estimating density. The notion of Effective Area Surveyed is formulated to allow use of all detections in estimating density.

The Emlen method arises when a particular view of the detectability curve is held. Other views lead to other methods. The Emlen method has the kind of flexibility best able to deal with the particular problems of surveying birds.

To improve its efficiency, we present a modernized version of Emlen's method based on analysis of a Cumulative Detection Curve.

Previous papers in this symposium dealing with the methods of estimating population density have given little, none, or disparaging mention of a method first proposed by J. T. Emlen (1971). It has been termed inefficient, lacking in theoretical foundation, highly subjective and sensitive to arbitrary data groupings. Yet virtually every ornithological paper we have seen that actually attempts to estimate density—either from line transect or variable circular plot surveys—uses Emlen's method directly or in a modified form. Why? It has been suggested that the reason for this is that more efficient methods have not previously entered the ornithological literature. True as that may be, and welcome as the newer methods should be, the purpose of this paper is to demonstrate that Emlen's method need not, as a result, be discarded. It is quite possible to modify Emlen's method to counter criticisms while maintaining its conceptual framework. This, for the most part, we attempt here.

In the following sections we present the conceptual framework of Emlen's method; examine the coefficient of detectability yields to the effective area surveyed as a measure of sampling effort; describe a design for the data analysis of a large survey; and describe a general graphical method for interpreting results of line transect and variable circular plot surveys alike.

EMLEN'S METHOD AND THE CD

Perhaps the least understood feature of Emlen's method is the coefficient of detectability, or CD. Yet the CD, and its cousin—the Effective Area Surveyed (EAS), play an indispensible role in the data analysis.

Let us begin with an abstract view, as in Figure 1 below, of a target region $\mathcal{S}$ of habitat over which a species has uniform density D.

What this means is that the average number of birds, $\mathscr{E}(m)$, to be expected in any specific subregion $\mathscr{R}$ with area $A(\mathscr{R})$ is $\mathscr{E}(m) = D \cdot A(\mathscr{R})$. In particular, if N is the number of birds in the entire region, then we have

$$D = \mathscr{E}(m)/A(\mathscr{R}) = \mathscr{E}(N)/A(\mathcal{S}) \qquad (2.1)$$

Now place an observer at the point "O," say, counting birds. Suppose there are n birds detected, m of which are in the particular region $\mathscr{R}$. If we suppose there is perfect detectability in $\mathscr{R}$, then m is all the birds present in $\mathscr{R}$, so that $m/A(\mathscr{R})$ unbiasedly estimates D. Furthermore, $\hat{N} = [m/A(\mathscr{R})] \cdot A(\mathcal{S})$ unbiasedly predicts the number N in the entire region, in the sense that N and $\hat{N}$ have the same average.

What is commonly understood to be Emlen's method consists of the following steps: (1) determine, from detection distance data, a region $\mathscr{R}$ of perfect detectability (we refer this to a *basal* region.); (2) estimate density in the target region by the observed density in the basal region; and (3) calculate the coefficient of detectability as CD = $n/\hat{N}$.

Stated in this way, the method resembles closely that of Kelker (1945), with the exception that the basal region is predetermined by Kelker and determined from the data by Emlen. The criticisms leveled at the method are these: (1) the n-m birds detected outside the basal region are not used to estimate density, except insofar as their locations help determine $\mathscr{R}$; (2) the CD is influenced strongly by the limits of the target region, yet the limits are essential; (3) the method uses grouped data and is therefore sensitive to the grouping procedure; (4) the density estimate is not statistically efficient (Burnham et al. 1980); and (5) what does one do with a CD?

Defense of Emlen's method is based on clarifying several points. First, we replace the CD by an effective area measurement, the EAS.

[1] Department of Statistics, Oregon State University, Corvallis, Oregon 97331.

[2] U.S. Fish and Wildlife Service, Patuxent Wildlife Research Center, P.O. Box 44, Hawaii National Park, Hawaii 96718.

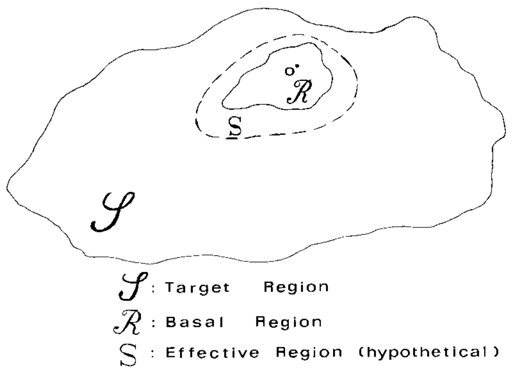

FIGURE 1. Conceptual Regions of an Emlen (Point) Survey.

Second we argue that this is not the proper time to estimate density. Then we present a simple, graphical technique for determining an estimate of EAS which does not require data grouping.

THE CD AND THE EAS

To begin, $n/\hat{N}$ estimates the probability of the observer's detecting a single bird positioned randomly in $\mathscr{S}$. Make $\mathscr{S}$ bigger and this probability must decrease. Make $\mathscr{S}$ unbounded, as several authors have done, and the probability is theoretically zero. To see what should be estimated, irrespective of the limits of the target region, consider the total number, n, of detections, which must lie between m and N. Hence its expectation lies between that of m and N, producing this result.

$$D \cdot A(\mathscr{R}) = \mathscr{E}(m) \leq \mathscr{E}(n) \leq \mathscr{E}(N) = D \cdot A(\mathscr{S})$$

Writing $\mathscr{E}(n) = D \cdot \mathscr{A}$, it is clear that $\mathscr{A}$ must be an area measurement intermediate between the basal area and the target area. It is an area representative of the observer's total survey effort, and is therefore defined to be the *Effective Area Surveyed* (EAS), (see Ramsey and Scott 1979, or Ramsey 1979). Thus

$$\mathscr{A} = \text{EAS} = E(n)/D, \tag{3.1}$$

which, if known, allows us to estimate density unbiasedly from all detections with $\tilde{D} = n/\text{EAS}$. It is sometimes theoretically convenient to view the EAS as the area of an effective *region* surveyed and to treat the whole procedure as one where the effective region is fully covered by the observer, while nothing is recorded outside of it. Such a region, S in Figure 1, is only, however, a hypothetical construct and perhaps should be deemphasized because of possible confusion with $\mathscr{R}$. At all costs, *avoid viewing EAS as the area of the basal region.*

Returning to (3.1), and incorporating (2.1), we get

$$\widehat{\text{EAS}} = \frac{\mathscr{E}(n)}{\mathscr{E}(m)} \cdot A(\mathscr{R}) \tag{3.2}$$

Theoretically, the EAS has this relation to the CD:

$$\text{CD} = \text{EAS}/A(\mathscr{S}).$$

The point here is that the EAS remains meaningful as $A(\mathscr{S})$ increases, whereas the CD does not.

ORGANIZING THE DATA ANALYSIS

Emlen argued that the CD should have some universality, being the same in regions of differing bird densities. By combining information from various sources, better estimates of CD's can be obtained. Here is how this works in an analysis of the results of a survey, except that we use the EAS.

The analysis proceeds in phases. In Phase 1, divide the target region into subregions according to a scale of detectability. At one end of the scale lies open grasslands. At the other lies dense forest with a high, closed canopy and thick understory. Between the extremes are classes reflecting how well one expects to detect birds visually and vocally. Lump together as a set all detections of a particular species in a particular detection class made by a particular observer. Further subdivisions should be made on factors such as time of day, weather, etc., which affect detectability, if these are not uniform during the survey.

In Phase 2 of the analysis, consider each set separately, producing with each a *detection curve,* such as in Figure 2. Here the density of observed detections is plotted against distance from the observer. Then comes this version of Emlen's method.

(I) Determine, from examination of the detection curve, a basal region $\mathscr{R}$ of near perfect detectability.

(II) Estimate the effective area surveyed by—see (3.2)—

$$\widehat{\text{EAS}} = \left(\frac{n}{m}\right) \cdot A(\mathscr{R}). \tag{4.1}$$

We emphasize here that *the purpose of looking at a detection curve is to estimate EAS, not density.* It should also be noted that one may use whatever auxiliary information one has available to judge what should be a suitable basal region. For example, a species which is attracted to the observer should not be allowed a basal region including only the area near the observer.

Phase 2 will produce estimates of EAS in many, but not all sets. A procedure for smoothing and filling in the missing EAS values is outlined in Ramsey and Scott (1979). It involves a weighted least-squares regression of the available values of $\log(\widehat{EAS})$ on variables indicating detectability class and observer. The fitted model is used to produce a full array of $\widehat{EAS}$ values for each observer in each habitat class, the whole procedure being done for each species. The full value of such a procedure is apparent when one realizes that this often gives EAS estimates in sets which began with very few or even no detections. Similarly, with rarer species and few detections, we are still able to use similarity with other, more abundant species to estimate EAS values. We have found, as Emlen suggested, that observer effects and detectability class effects are quite consistent from species to species.

Phase 3 consists of estimating population density. Suppose we wish to estimate the average density of some species in a given subregion of the target region. Divide the subregion as before into detectability strata, according to observer and detectability class. Determine, in each stratum, the total area (A_j) it occupies in the subregion, the total area (a_j) effectively surveyed, and the total number (n_j) of detections. The latter two are found by summing over pieces of transects or over stations, depending on how the survey was conducted. Then estimate the average density in the subregion to be

$$\hat{D} = \left(\sum_j n_j A_j / a_j\right) \Big/ \left(\sum_j A_j\right).$$

One expression which estimates the variability in $\hat{D}$ is

$$\widetilde{\mathrm{Var}}(\hat{D}) = \left[\sum_j n_j (A_j/a_j)^2\right] \Big/ \left[\sum_j A_j\right]^2.$$

This treats the effective areas as having been estimated without error and treats the numbers present as variables. Modifications may be made to recognize errors in EAS estimates. And in certain (management) situations, it may be preferable to hold fixed the numbers present.

SELECTING THE BASAL REGION

When J. T. Emlen (1971) proposed his method, he suggested that the basal region be found by inspecting the detection curve for a point of inflection, where density begins to decline rapidly with distance. Ramsey and Scott (1979) discussed several ways to formulate a rule that would replace "inspection" and settled on a scheme which uses likelihood ratios to judge if density is declining.

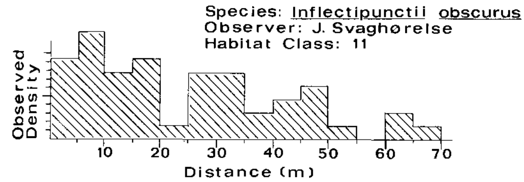

FIGURE 2. A plot of the density of detections versus the distance of detection for a hypothetical species and observer.

We emphasize that the purpose for devising such a rule was NOT simply to facilitate automatic data processing in a high speed computer. Detection curves should always be plotted and visually inspected. Only in this way can one understand the factors influencing detectability. The reason for the rule was to provide a method less subject to influence of random variations.

The likelihood ratio rule says that a basal region $\mathcal{R}$ should be expanded to include $\mathcal{R}^*$ if a statistical test finds no difference in density in the two regions. It incorporates a flexible critical ratio which may be chosen to provide balance between bias and variability in the resulting estimators. We choose a "conservative" cutoff value which usually underestimates density by 10–15% (see DeSante 1981), because this greatly reduces the possibility of seriously over-estimating density. (Our primary concern is with rare and endangered species, whose population sizes we do not want to over-estimate).

RELATED METHODS

There are a number of ways to estimate Effective Area Surveyed from detection distance data, Emlen's method being just one. Burnham et al. (1981) argue that the EAS bears a known relationship with the probability density function of detection distances (in line transects, but *squared* distances in circular plots), evaluated at zero distance. Ramsey (1979) suggests incorporating the EAS as a scale parameter in a flexible family of possible detectability curves. The choice of procedure here depends largely upon how one feels about the detectability curve. If one feels confident that detectability curves belong to a certain parametric family, then Ramsey's (1979) methods provide highly efficient estimators. If one is confident that all birds on transect (station) are detected but that detectability declines rapidly off transect (station), then the Burnham et al. (1981), non-parametric procedures might be best. However, if one feels

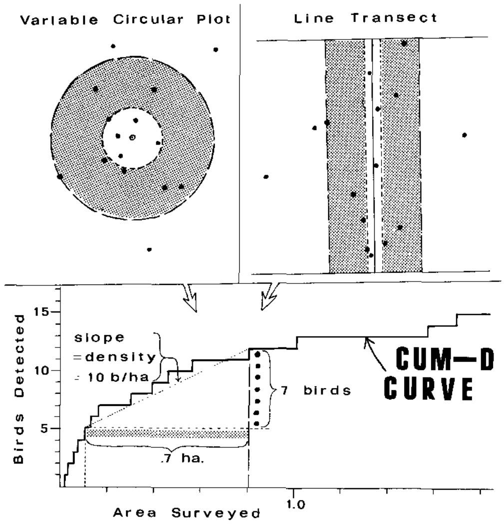

FIGURE 3. In this figure, a VCP and a LT survey (upper left and right, respectively) give the same CUM-D curve. Against Area on the abscissa, plot the number of detections made in that area around the observer. The slope through a part of the curve then gives the density of detections over the corresponding (shaded) region.

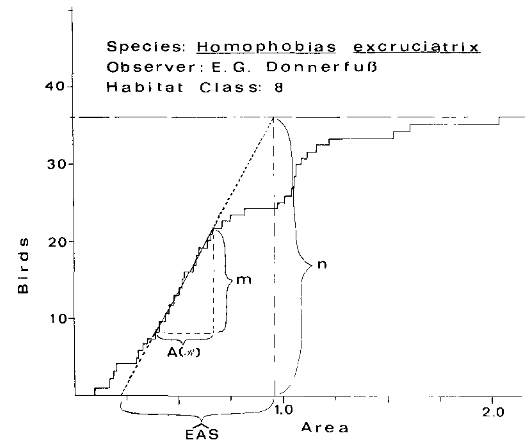

FIGURE 4. To estimate EAS, extend the maximum CUM-D curve slope up to the horizontal line of total detections, and extend it down to the horizontal axis. The $\widehat{EAS}$ is the area difference between these points.

that there is some substantial region of uniform, near-perfect detectability, the modified Emlen technique is recommended.

EMLEN'S METHOD WITHOUT GROUPING

In this section we introduce a function which can be used to apply Emlen's method graphically to estimate EAS. The function is the CUMulative Detection Curve (CUM-D), which displays total numbers of detections as a function of area searched. It is applicable to both Line Transect surveys and to Variable Circular Plot surveys, as illustrated in Figure 3, where the two survey results at the top (dots represent detections) produce identical CUM-D curves. Specifically, we break up the region surveyed into zones of increasing area surrounding the observer's position(s). The CUM-D curve plots the total number of detections in a zone against the area of that zone. From the CUM-D curve, one may calculate the density of detected birds in any subzone. For example, the shaded regions have the same area, 0.7 ha, in Figure 3 and have the same number, 7, of detected birds. Therefore, the density of detections in the shaded region is—as illustrated—the slope of the CUM-D curve between the inside and outside areas.

Statisticians will tell you that division of the CUM-D curve by n will produce the "empirical distribution function" of the areas, $A_1, \ldots, A_n$, enclosed by the detections. Here

$$A_j = \begin{cases} \pi \cdot R_j^2, \text{ in a VCP, where } R_j = j^{\text{th}} \text{ detection distance} \\ 2 \cdot L \cdot Z_j, \text{ in a } LT \text{ of length } L, \text{ where } Z_j = j^{\text{th}} \text{ right-angle distance.} \end{cases}$$

Note that the slope of the CUM-D curve at $A = 0$ is the critical parameter estimated by Burnham et al. (1981).

Because density of detections is highest in zones of highest detectability, we now offer a final version of Emlen's method, to wit:

(1) Determine a basal region $\mathcal{R}$ by seeking a zone of highest slope in the CUM-D curve.
(2) Estimate EAS by projecting the slope to "0" and "n" detections, as in Figure 4. (Equivalent to Equation (4.1).)
(3) Use EAS values (smoothed, if appropriate) to estimate densities.

Methods for selecting a largest slope and resultant properties of estimators are discussed in Wildman (pers. comm.).

DISCUSSION

There is no requirement that a basal region include the zone immediately surrounding the observer. In bird surveys, investigators have encountered observer avoidance problems in variable circular plot surveys where it might not be anticipated and in line transect surveys where it is to be expected. Indeed, it is almost inconceivable that birds would not react to the presence of an observer. This invalidates the assumption that detectability at zero distance is perfect ($g(0) = 1$). It need *not* preclude the determination of accurate estimates of population density.

The modified Emlen technique produces a basal region wherein observed densities are highest. Once obtained, the investigator still must relate that observed density to population density. Are birds moving away from the observer and then resuming normal behavior? Are birds near the observer simply making themselves undetectable? If so, how does the zone of avoidance compare with the zone of high detection? Is some fraction ($p = ½$, say, for females) of the population completely non-detectable? Are the birds moving over considerable distances in the time they are observable? Are distances misjudged because of ventriloquism? These are some of the questions one must confront in relating observed to true densities. Although the full shape of the CUM-D curve or the detection curve can be useful in the discussion, answers must, in the final analysis, be based largely on the biology of the target species.

There will be species for which these methods fail totally. But, there will also be those for which it works.