The Condor 92:313–325

ANALYST AND OBSERVER VARIABILITY IN DENSITY ESTIMATES FROM SPOT MAPPING[1]

JARED VERNER AND KATHLEEN A. MILNE[2]
USDA Forest Service, Pacific Southwest Forest and Range Experiment Station, 2081 E. Sierra Ave., Fresno, CA 93710

Abstract. We studied variation in density estimates from spot mapping that was attributable to analysts and observers, using expert birders with little or no prior experience with spot mapping. Three observers independently spot mapped one 42-ha plot (Markwood) in mixed-conifer forest, and four others independently spot mapped another (Teakettle). All observers analyzed all maps. Consistency among analysts and observers in estimating the numbers of territories of breeding species on each plot was generally poor. Across all combinations of analysts and maps, 71% of all ANOVAs had significant analyst and/or observer effects. Observer effects were generally greater than analyst effects. When observers analyzed their own species maps, CVs of individual species ranged from 0% to 173% (mean 41%) at Markwood and from 0% to 188% (mean 60%) at Teakettle. As in similar studies, mean CVs from pooled totals of all species were less than those from individual species and were within the range of variation found by other researchers. Based on the range of CVs observed among species in this study, the number of sample plots needed to detect a statistical difference in density of a given species between samples is probably prohibitive for most studies. Instead, practitioners need to design studies to control observer and analyst variability.

Key words: Spot mapping; observer variability; analyst variability; coefficient of variation; sampling design.

INTRODUCTION

Ornithologists generally consider density estimates based on spot mapping to be more accurate than those from any other method. Indeed, many researchers have used results from spot mapping as their standard for comparing density estimates from other methods (e.g., Howell 1951; Stewart et al. 1952; Enemar 1959; Emlen 1971, 1977; Franzreb 1976, 1981; Dickson 1978; Järvinen et al. 1978a, 1978b; DeSante 1981, 1986; Hildén 1981; O'Meara 1981; Hamel 1984; Verner and Ritter 1988). Although many of these workers have acknowledged that even spot mapping may not always deliver accurate density estimates, few studies have explored sources of variation in the method and the impact of that variation on density estimates. As Enemar et al. (1978, p. 38) pointed out, "Very few studies on observer variability have been carried out although this problem is central for determining the accuracy of a comparison between census results from different areas and different time periods. Of course, the problem is most significant in small, specialized studies with few census workers." The scant literature on this topic reports variations attributable only to analysts and/or observers (Snow 1965; Hogstad 1966; Bell et al. 1968, 1973; Svensson 1974; Best 1975; Enemar et al. 1978; O'Connor 1981, O'Connor and Marchant 1981). However, many of those studies were ad hoc, having been designed with other objectives in mind, so efforts to distinguish analyst and/or observer effects were confounded by other sources of variation.

Our objectives were (1) to quantify the magnitude of variation in density estimates from spot mapping attributable to analysts and observers, and (2) to estimate sample sizes (number of plots) needed to compare with known confidence the density estimates of given bird species between habitat types.

STUDY AREA

Two 42-ha plots, "Markwood" and "Teakettle," were established in the west-central Sierra Nevada, Fresno County, California. Markwood had interspersed patches of mature and old-growth mixed-conifer forest (not logged since selective cutting in the late 1960s), montane chapparal, and granitic outcrops with scattered shrubs and

[1] Received 9 June 1989. Final acceptance 26 December 1989.
[2] Current address: Coronado National Forest, P.O. Box 709, Safford, AZ 85548.

black oaks (*Quercus kelloggii*). Two small meadows and three small streams added habitat diversity. Mean tree-canopy cover was 52%, based on vertical projections from 169 points regularly spaced in a grid pattern at 50-m intervals throughout the plot. Dominant canopy trees were white fir (*Abies concolor*), incense cedar (*Libocedrus decurrens*), and sugar pine (*Pinus lambertiana*). Mean shrub cover was 23%, based on the same points used to estimate tree-canopy cover. Dominant shrub species were mountain whitethorn (*Ceanothus cordulatus*), hazelnut (*Corylus cornuta*), snowberry (*Symphoricarpos* sp.), Sierra gooseberry (*Ribes roezlii*), Sierra currant (*Ribes nevadense*), and greenleaf manzanita (*Arctostaphylos patula*). Bracken fern (*Pteridium aquilinum*) was an abundant and widespread groundcover.

Teakettle was in an old-growth mixed-conifer forest that had been selectively logged only for road-building purposes and to remove trees considered hazardous near cabins or roadways. Its periphery was primarily dense forest, but the forested center of the plot also included patches of shrubs, rock outcrops, and occasional black oaks. As at Markwood, this plot included three small streams and two small meadows. Tree-canopy cover averaged 47%. Dominant tree species were white fir, red fir (*Abies magnifica*), sugar pine, Jeffrey pine (*Pinus jeffreyi*), and incense cedar. Mean shrub cover was 17%. Dominant shrub species were mountain whitethorn, snowberry, hazelnut, chinquapin (*Castanopsis sempervirens*), Sierra gooseberry, and greenleaf manzanita.

METHODS

BIRDS

Densities of territorial birds were estimated using our version of spot mapping, based largely on the international standard (Anonymous 1970). Square mapping plots, 650 m on each side, were established with grid lines at 50-m intervals. A narrow path along each line was cleared of shrubs and fallen woody material to facilitate walking by observers. Steel fence posts with alphanumeric codes marked each grid intersection. Three observers independently sampled the Markwood plot and four independently sampled the Teakettle plot. Each observer completed 12 visits to a plot, beginning at 05:00 and ending by 14:00 PST, from 20 May to 12 June 1986. Observers were constrained from discussing any field observations among themselves until all individual species maps had been interpreted by all observers.

Each visit involved walking and mapping seven alternate grid lines so that the observer passed within 50 m of every point on the plot. The initial line and direction walked were randomly chosen for each visit, with the constraints that each observer initiated visits three times along each side of the grid and each grid line was walked three times during a 12-visit effort. Bird detections were recorded on separate daily field sheets for each line walked during each visit. Observers endeavored to avoid registering the same bird more than once while walking a given line; however, no effort was made to avoid duplicate registrations of individuals when walking different lines. Registrations were transferred to species maps using different letters to designate visits and different colors to designate lines walked during a given visit. In this way, duplicate registrations of the same individual from different lines on the same visit could be identified and appropriately interpreted from the complete record on the species maps. Fractions of boundary territories assigned to a plot were either (1) the proportion of registrations for a given boundary cluster that were within the plot, for species with fewer than three complete clusters, or (2) the number of registrations in boundary clusters that were within the plot divided by the mean number of registrations in complete clusters, for species with at least three complete clusters.

The study design controlled effects of weather, day, season, year, and habitat on density estimates. (All observers sampled the same plots in fair weather on the same days in the same year, following a randomized series of assignments to starting points and directions of movement over the plots.)

Most observers had no prior experience with spot mapping or with the study plot to which they were assigned. At Markwood, one observer had spot mapped and analyzed species maps during two prior seasons, one at Markwood. Another had one season's prior experience at Markwood, and the third had no prior experience with spot mapping or with Markwood. At Teakettle, one observer had one season's prior experience with spot mapping and analysis of species maps but no prior experience with the Teakettle plot. The other three observers had no prior experience with spot mapping or with Teakettle.

O'Connor (1981) reported that naive analysts do not adequately follow standardized instructions of the International Bird Census Committee or the British Trust for Ornithology. Because we had observed the same tendency among inexperienced analysts in previous years of our studies, we implemented training sessions for both field and office phases of spot mapping. Observers first spent 1 day in training to analyze species maps, using maps of one or two species obtained at Markwood and Teakettle in an earlier year. This included extensive discussions of all rules, and the importance of adhering to them. They then spent nearly 2 weeks in the field renewing their skills with identification of songs and calls of birds resident in forest habitat similar to that on the plot they would sample. Observers then spent 1 day spot mapping on a plot other than the one they would sample, and results of the day's field effort were discussed among all observers. Finally, they spent 2 days on the actual mapping plot, clearing and flagging trails, and marking grid intersections.

ANALYSES

Each observer first analyzed species maps based on their own field effort. Then each analyzed species maps based on field efforts of all other observers. Species maps were analyzed using acetate overlays and results were transferred to photocopies of the species maps. In this way, all analysts could use the same, original species maps, and density estimates could be compared in four cases: Case 1—observers analyzed their own species maps; Case 2—analysts interpreted species maps of other observers who had worked on the same plot that they had; Case 3—analysts interpreted species maps of other observers who had worked on a different plot from the analysts; and Case 4—all analysts interpreted species maps, irrespective of their source.

The total numbers of territories of all species pooled were determined for each combination of analyst and observer (49 possible combinations). These totals were then used in analyses of variance (ANOVAs) to test for differences among analysts and among observers when determining the total number of territories, using PROC ANOVA from SAS (SAS Institute 1982). The model for Case 1, above, was

$$T_i = u + e_i \qquad (1)$$

where T_i was the total number of territories for observer-analyst i, u was the overall mean number of territories, and e_i was the error term. The model for Cases 2, 3, and 4, above, was

$$T_{ij} = u + o_i + a_j + e_{ij} \qquad (2)$$

where T_{ij} was the total number of territories for observer i and analyst j, u was the overall mean number of territories, o_i was the effect due to observer i, a_j was the effect due to analyst j, and e_{ij} was the error term.

A nonzero interaction term (the interaction between analyst and observer) may have been present, but our study design did not allow us to estimate such a possibility. By leaving out that term, we have assumed that it was small compared to the main effects and did not significantly inflate the error term. In addition, because the same individuals analyzed three or four sets of maps of each species from the same plots, we violated the assumption of independent data when using ANOVA. The extent to which this affected results is unknown.

ANOVAs of individual species' territory totals (only for species with two or more territories), as derived from each analyst/observer pair, were computed, using the same models as used for Case 1 and for Cases 2, 3, and 4.

We used the coefficient of variation (CV) as a measure of variability across analysts and observers, because it gives a standardized value that allows comparison among species. The formula used for CV,

$$CV = SD/\bar{x} \qquad (3)$$

follows the accepted statistical definition. Because other workers have computed CVs differently, or did not compute them at all, we did so from their published data when possible, using Formula 3 to standardize their results with ours. CVs of small samples (n = 3 or 4 in our case) tend to be negatively biased. Because this bias is too small to alter conclusions from this study, we made no bias adjustments. Tests of statistical significance are based on an alpha level of 0.05.

RESULTS

RANK ORDER OF ABUNDANCE

The number of species for which territories were detected ranged from 41 to 42 ($\bar{x}$ = 42; SD = 0.58; CV = 1.4%) at Markwood and from 31 to 36 ($\bar{x}$ = 32; SD = 2.71; CV = 8.5%) at Teakettle, counting only species for which observers estimated at least a fractional territory on the plots (Table 1). The species were consistently ranked in about the same order of abundance for Case 1 (observers analyzed their own species maps). Spearman's rank-order correlations for all species

TABLE 1. Estimated numbers of territories for each species on the two study plots for Case 1 (observers analyzed their own species maps); + indicates presence on the plot but too few registrations to confirm a territory.

Species	Markwood observers			Teakettle observers			
	A	B	C	D	E	F	G
Sharp-shinned Hawk (*Accipiter striatus*)	+	+	+	+	+	+	+
Cooper's Hawk (*Accipiter cooperii*)	0.9	1.0	+	+	+	+	+
Red-tailed Hawk (*Buteo jamaicensis*)	+	+	+	+			1.0
Mountain Quail (*Oreortyx pictus*)	6.3	5.3	10.5	2.5	2.5	2.3	2.0
Band-tailed Pigeon (*Columba fasciata*)	+	+	1.0		+		+
Northern Pygmy-Owl (*Glaucidium gnoma*)				+	1.0		+
Spotted Owl (*Strix occidentalis*)	+	+					
Northern Saw-whet Owl (*Aegolius acadicus*)	+						
Common Nighthawk (*Chordeiles minor*)	+		+				
Calliope Hummingbird (*Stellula calliope*)	+	+					
Red-breasted Sapsucker (*Sphyrapicus ruber*)	1.0	1.8	2.7	4.3	3.4	2.3	3.8
Hairy Woodpecker (*Picoides villosus*)				2.7	2.9	1.7	1.9
White-headed Woodpecker (*Picoides albolarvatus*)	4.2	4.4	4.6	5.1	4.1	4.1	3.6
Northern Flicker (*Colaptes auratus*)	0.8	0.8	0.8	0.8	0.9	0.8	0.4
Pileated Woodpecker (*Dryocopus pileatus*)	0.9	0.9	0.8	0.9	1.6	1.0	0.8
Olive-sided Flycatcher (*Contopus borealis*)	+	+			1.5		
Western Wood-Pewee (*Contopus sordidulus*)	9.2	9.8	7.4	1.6	1.7	1.5	1.7
Hammond's Flycatcher (*Empidonax hammondii*)	26.4	20.9	38.7	16.9	23.8	13.8	34.1
Dusky Flycatcher (*Empidonax oberholseri*)	9.6	10.8	10.4	16.9	20.9	15.9	17.0
Western Flycatcher (*Empidonax difficilis*)	3.0	3.0	3.3				
Steller's Jay (*Cyanocitta stelleri*)	0.9	2.2	9.3	2.7	3.5	3.1	3.5
Common Raven (*Corvus corax*)	0.9	0.9	+	0.9	0.9	0.9	0.9
Mountain Chickadee (*Parus gambeli*)	7.2	15.2	11.9	8.8	9.1	9.6	8.9
Red-breasted Nuthatch (*Sitta canadensis*)	11.7	21.0	26.9	12.2	28.4	14.2	11.7
White-breasted Nuthatch (*Sitta carolinensis*)	1.0	+	+				
Brown Creeper (*Certhia americana*)	6.0	10.9	10.0	5.3	12.3	7.0	5.0
House Wren (*Troglodytes aedon*)	+	1.0	1.0				
Winter Wren (*Troglodytes troglodytes*)	1.1	2.0	4.2	4.7	12.4	4.0	3.9
Golden-crowned Kinglet (*Regulus satrapa*)	27.4	38.1	29.7	22.7	41.0	15.8	25.4

TABLE 1. Continued.

Species	Markwood observers			Teakettle observers			
	A	B	C	D	E	F	G
Townsend's Solitaire (*Myadestes townsendi*)	1.8	0.6	0.7	3.5	3.4	4.2	3.9
Hermit Thrush (*Catharus guttatus*)	5.8	4.9	7.3	9.3	4.0	4.0	6.0
American Robin (*Turdus migratorius*)	5.9	9.9	10.4	1.0	3.6	1.7	1.0
Solitary Vireo (*Vireo solitarius*)	13.9	16.9	16.8	3.4	1.0	1.6	1.5
Warbling Vireo (*Vireo gilvus*)	10.9	11.6	18.9	2.7	3.3	0.8	1.3
Nashville Warbler (*Vermivora ruficapilla*)	6.9	5.5	12.0	10.2	17.8	9.3	9.0
Yellow Warbler (*Dendroica petechia*)	2.7	2.9	5.0				
Yellow-rumped Warbler (*Dendroica coronata*)	17.5	24.6	12.1	17.2	17.8	5.6	14.7
Hermit Warbler (*Dendroica occidentalis*)	33.1	25.8	44.6	19.8	62.2	13.8	39.2
MacGillivray's Warbler (*Oporornis tolmiei*)	11.2	23.6	17.8	13.2	6.1	8.4	6.7
Wilson's Warbler (*Wilsonia pusilla*)	2.3	2.7	6.5				
Western Tanager (*Piranga ludoviciana*)	10.9	10.5	13.7	11.9	6.8	13.7	15.1
Black-headed Grosbeak (*Pheucticus melanocephalus*)	3.9	4.8	6.2		1.8		
Green-tailed Towhee (*Pipilo chlorurus*)	0.9	3.0	1.0	2.0	1.0		+
Rufous-sided Towhee (*Pipilo erythrophthalmus*)	2.8	3.1	5.0				
Chipping Sparrow (*Spizella passerina*)	1.5	1.3	1.9		1.0		
Fox Sparrow (*Passerella iliaca*)	52.2	61.7	58.8	58.5	71.7	44.2	75.7
Song Sparrow (*Melospiza melodia*)	0.7	1.0	0.7				
Lincoln's Sparrow (*Melospiza lincolnii*)	1.7	2.9	1.9				
Dark-eyed Junco (*Junco hyemalis*)	19.2	28.3	35.0	20.4	55.5	19.3	20.7
Brewer's Blackbird (*Euphagus cyanocephalus*)		+					
Brown-headed Cowbird (*Molothrus ater*)	+	+	0.8		+	+	
Purple Finch (*Carpodacus purpureus*)	10.5	9.6	11.8				
Cassin's Finch (*Carpodacus cassinii*)	0.9	+	1.0	+	1.7	+	+
Pine Siskin (*Carduelis pinus*)	0.3	0.5	0.5		+		
Lesser Goldfinch (*Carduelis psaltria*)	+	0.8					
Evening Grosbeak (*Coccothraustes vespertinus*)	1.0	0.8	+	1.9	2.5	0.8	+
Species richness							
Species with at least a fractional territory on the plot	42	42	41	31	36	31	30
Total, including species recorded only as present (+)	53	52	48	36	41	35	37

TABLE 2. Mean CV ± SE and median CV of density estimates for three subsets of species on each plot, based on cases in which observers analyzed their own species maps: (1) all species with at least a fraction of one territory estimated by at least one observer, (2) species for which all observers estimated at least a fraction of one territory, and (3) species with two or more territories estimated by each observer.

Plot	All species	Species for which all observers had a territory value	Species with two or more territories
Markwood			
Mean	44.6 ± 6.51 ($n = 45$)	30.5 ± 3.63 ($n = 37$)	26.3 ± 2.55 ($n = 26$)
Median	32.0	29.0	28.5
Teakettle			
Mean	58.9 ± 9.68 ($n = 35$)	33.7 ± 3.84 ($n = 28$)	32.4 ± 4.24 ($n = 20$)
Median	41.0	34.5	36.8

on each plot ranged from 0.91 to 0.97 ($P \ll 0.01$) for all pairwise comparisons of observers (three cases at Markwood; six cases at Teakettle). When only the 10 most abundant species in each observer's rank order were compared in all pairwise cases, correlation coefficients were considerably lower (0.57–0.82), but all were significant—two at the 0.05 level and seven at the 0.01 level. The Fox Sparrow (*Passerella iliaca*) was ranked as the most abundant species by all observers on each plot, and only 12 species at each plot included the 10 most abundant for all observers combined.

NUMBER OF TERRITORIES: OBSERVER EFFECTS

Unlike the case for rank order of abundance, consistency in estimating the numbers of territories of various species on each plot was generally poor for Case 1 (Table 1). All persons had identical density estimates for only one species at each plot—Northern Flickers (*Colaptes auratus*) at Markwood and Common Ravens (*Corvus corax*) at Teakettle. Differences were the rule, and often they were extreme. For example, for Case 1, using only species with at least a partial territory, the highest estimate by any person was at least twice the lowest by any person for 11 species (30%) at Markwood and 16 species (57%) at Teakettle. The mean ratio of the highest to the lowest estimate was 2.03 at Markwood (SD = 1.54; range = 1–10.3; $n = 37$) and 2.24 at Teakettle (SD = 0.94; range = 1–4.5; $n = 28$). The extreme case was that of the Steller's Jay (*Cyanocitta stelleri*) at Markwood (ratio of high to low = 10.3). All observers had about the same number of registrations for this species (70, 75, and 75), but one had several more territorially significant registrations than the others. And one observer was unusually liberal when applying criteria for delineating clusters, whereas another was unusually conservative.

CVs of the number of territories estimated for each species by observers at each plot, for Case 1, tended to be high, ranging from 0% to 175% ($\bar{x} = 44.6\%$) at Markwood and 0% to 188% ($\bar{x} = 58.9\%$) at Teakettle. Mean CVs were highest for all species combined, primarily because these included species for which not all observers recorded at least a partial territory on the plot. Mean CVs were lower for the subset of species for which all observers detected at least a partial territory, and they were lowest for even more common species—those for which each observer detected at least two territories on the plot (Table 2). Mean CVs were significantly higher ($P < 0.01$; *t*-test) at Teakettle than at Markwood for all three of the above groups of species (Table 2).

Median CVs in relation to the means indicated some skew in the data for all species on both plots and for species with two or more territories by each observer at Teakettle. However, medians and means were nearly the same on both plots for species with a territory value by all observers and for species with two or more territories by each observer at Markwood (Table 2).

NUMBER OF TERRITORIES: ANALYST EFFECTS

Twenty-six species at Markwood and 20 at Teakettle had at least two territories estimated by all observers. Mean CVs of these species, using (1) all analysts, (2) just Markwood observers as

TABLE 3. Mean CV ± SE of density estimates of different groups of analysts, using data only from species with an average of at least two territories, by plot and analyst group.

Plot	Analyst group		
	All observers	Markwood observers	Teakettle observers
Markwood	14.6 ± 1.21 (n = 26)	12.0 ± 1.14 (n = 26)	14.5 ± 1.21 (n = 26)
Teakettle	27.2 ± 1.92 (n = 20)	27.8 ± 3.29 (n = 20)	22.1 ± 1.89 (n = 20)

analysts, and (3) just Teakettle observers as analysts, were consistently less for the Markwood than the Teakettle species (Table 3) ($P < 0.01$ in all three cases; t-test). Mean CVs of the total territories of Markwood species were lowest when analyzed by Markwood observers, and those of Teakettle species were lowest when analyzed by Teakettle observers (Table 3), but differences were not significant (t-tests).

Using territory counts from all seven analysts, 18 of the 26 species at Markwood and all 20 at Teakettle had significant ANOVAs with respect to analyst, observer, or both. Seventy-one percent of all ANOVAs had significant analyst and/or observer effects (Table 4). Five percent had significant analyst effects only, 40% had significant observer effects only, and 26% had significant effects of both analyst and observer.

In most instances, density estimates for a given species were higher for Case 1 than for Cases 2 or 3. Using the subset of species for which each observer had at least two territories (the more common species), we found no evidence by ANOVA that results for Case 1 differed significantly from results for Cases 2 or 3. No species at Teakettle and only one at Markwood differed significantly between these cases. In that instance, the higher number of territories occurred for Case 1. On the other hand, 22 of the 26 more common species at Markwood had higher territory totals for Case 1 than for Cases 2 or 3 ($P < 0.001$, sign test); at Teakettle only 11 of 20 species had higher totals for Case 1 ($P = 0.41$, sign test). A similar comparison of all species showed that 37 of 42 species at Markwood had higher totals for Case 1 ($P < 0.001$, sign test), and at Teakettle it was 16 of 35 species ($P = 0.63$, sign test).

VARIATION WITH RESULTS POOLED ACROSS SPECIES

The number of territories of all species pooled, as determined by each analyst/observer pair, was highly variable (Table 5). ANOVAs revealed significant differences at both plots. Although results varied among the different groups of analysts—(1) Markwood observers analyzing Markwood maps, (2) Markwood observers analyzing Teakettle maps, (3) Teakettle observers analyzing Markwood maps, (4) Teakettle observers analyzing Teakettle maps, (5) all observers analyzing Markwood maps, and (6) all observers analyzing Teakettle maps—observer effects were generally more important than analyst effects (Table 6). The overall model was significant in five of the six cases, analyst effects were significant in three cases, and observer effects were significant in five cases. Analyst and observer effects were both significant in only two cases, and observer effects were greater than analyst effects in four cases.

Although pooled density estimates were higher

TABLE 4. Number of significant ANOVAs for individual species with two or more territories (26 species at Markwood, 20 species at Teakettle) by plot and by analyst groups.

Significant effects	Analysts of Markwood maps			Analysts of Teakettle maps			Total
	Markwood observers	Teakettle observers	All observers	Markwood observers	Teakettle observers	All observers	
Observer only	5	8	6	12	10	13	54
Analyst only	0	2	3	0	2	0	7
Both	4	3	9	4	7	8	35
Neither	15	11	6	5	2	0	39

TABLE 5. Total territories, pooled across species, for analysts and observers at both plots.[1]

	Markwood plot				Teakettle plot				
	Observers				Observers				
Analyst	A	B	C	$\bar{x}$	D	E	F	G	$\bar{x}$
A	321.6	342.3	391.2	351.7	202.8	534.9	288.1	309.5	333.8
B	339.6	389.4	350.4	359.8	217.0	388.7	244.6	237.3	271.9
C	411.2	423.3	443.4	426.0	222.3	669.1	305.9	265.9	365.8
D	290.4	338.1	278.3	302.3	273.1	441.7	277.4	297.0	322.3
E	325.0	346.1	288.5	319.9	259.1	415.9	242.2	249.8	291.8
F	261.2	344.7	260.4	288.8	188.5	335.0	221.5	225.5	242.6
G	291.2	353.3	313.6	319.4	236.7	506.7	266.7	314.3	331.1
$\bar{x}$	320.0	362.5	332.3	338.3	288.5	470.3	263.8	271.3	308.5

[1] Analyst A is the same person as observer A, analyst B is the same as observer B, and so on.

for Case 1 than for Cases 2 and 3 in five of the seven comparisons (Fig. 1), the tendency was not significant (Case 1 ANOVAs: $P = 0.10$ for Markwood, $P = 0.97$ for Teakettle).

DISCUSSION

VARIATION WITHIN SPECIES

Analyst effects. Although analyst effects were less than those of observers in our study, nonetheless they were often significant. Our mean CVs at Markwood were all lower than those found in other studies, but those at Teakettle were all higher (see Table 3). Svensson (1974) had 58 persons analyze the same 37 maps (thus observer effects were controlled) of six different species. The mean CV for all species maps and analysts was 21%. In a different trial reported in the same study, 17 other persons analyzing six species maps had a mean CV of 20%. Best (1975) gave copies of the map for one species on one plot to five analysts; the CV of their territory totals was 23%. Because Best's was a color-banded population, he was able to determine that all analysts underestimated the true population. O'Connor and Marchant (1981) found significant analyst effects for only two of the 26 species with at least two territories on their plot. It is important to note that analysts in the latter study had considerable experience and in-depth training to prevent individual idiosyncrasies of analysis from influencing density estimates.

Observer effects. Most previous studies have not reported the extent of analyst and observer effects on intraspecific variation in estimates of density. O'Connor and Marchant (1981) found significant observer effects for 24 of the 26 species that had at least two territories on their plot. The median CV was about 52% for observer effects, and we calculated a mean CV of 52% for all 45 species reported on their plot. This compares favorably with our mean CVs for all species at Markwood (41%) and Teakettle (60%). Similarly, our results were within the range of CVs computed for other studies, using only those species for which all analysts identified territories on the respective study plots (Table 7). Results of our study, and those of O'Connor and Marchant (1981), suggest that analyzing only a subset of more common species gives a markedly lower mean CV and, thus, a more optimistic picture of observer variability than is the real case (e.g., see Table 2; also the CV for more common species in O'Connor and Marchant's study in Table 7).

VARIATION WITH RESULTS POOLED ACROSS SPECIES

Analyst effects. All available information suggests that analyst effects are less than those of observers. For example, the mean CV across analysts in the study by O'Connor and Marchant was only 1%. In our study it was 14% at both

TABLE 6. ANOVA P-values for the total number of territories with respect to observer and analyst, by analyst groups.

Analysts	Source of variance	Markwood maps	Teakettle maps
Markwood observers	Model	0.0622	0.0090
	Analyst	0.0325	0.2025
	Observer	0.2414	0.0042
Teakettle observers	Model	0.0126	0.0001
	Analyst	0.1577	0.0063
	Observer	0.0038	0.0001
All observers	Model	0.0007	0.0001
	Analyst	0.0006	0.0574
	Observer	0.0276	0.0001

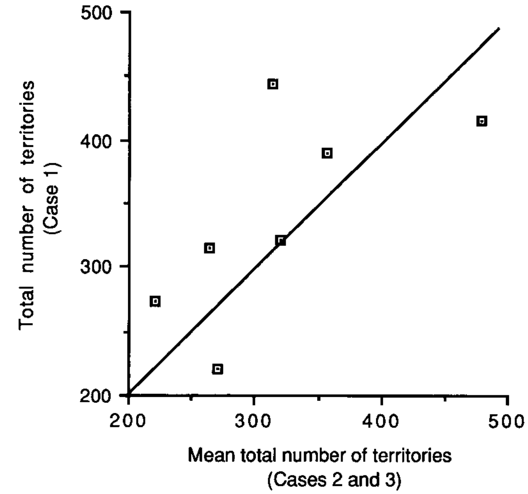

FIGURE 1. Total number of territories for Case 1 (observers analyzed their own species maps) vs. the corresponding means of Cases 2 and 3 (analysts interpreted maps of other observers). Instances when observers interpreted more territories from their own maps than the mean of all other analysts interpreting the same maps fall above the diagonal; the reverse was true of points below the diagonal.

Markwood and Teakettle. Again, extensive analyst training in the case of O'Connor and Marchant's study was probably a factor. We have not been able to extract similar values for data pooled across species in other studies, although Svensson's (1974) study using 58 different analysts produced a range in number of estimated territories from 191 to 425 among the 37 maps analyzed. The comparable maximum range in O'Connor's (1981) study was from 239 to 255, and in our study it was from 260 to 443 at Markwood and 335 to 669 at Teakettle.

Observer effects. In all but one instance (Teakettle, this study), CVs from results pooled across species were less than the mean CVs from individual species. Our pooled results, with the exception of observer effects at Teakettle, were well within the range of variation found by others (Table 8). Exact comparisons are complicated, however, because the designs of all studies differed in one or more meaningful ways. For example, CVs for Enemar (1962), Hogstad (1966), and Enemar et al. (1978) were based on total detections of each species by different observers, rather than the number of territories estimated. Snow (1965) and O'Connor and Marchant (1981) mixed results from single observers with those from teams of two observers. In O'Connor and Marchant's (1981) study, only one of the four observers analyzed their own maps. Final map interpretations by Enemar et al. (1978) were based on a consensus of the four observers, thus negating any chance of obtaining independent estimates among analysts. Finally, the level of experience among analysts and observers differed markedly, probably with those in our study having the least experience.

OBSERVER AND ANALYST EXPERIENCE

The experience of analysts and observers did not appear to affect overall results in our study, as CVs of our pooled data were comparable to those of other studies in which both observers and analysts were highly experienced. Similarly, Svensson (1974, p. 331–334) found "no clear indication that differences exist which can be explained by the degree of experience with the mapping method." On the other hand, O'Connor (1981, p. 374) indicated that "extensive experience of the mapping method . . . can lead to an improved detection of breeding pairs on the plot." We suspect that this is true and that the similarity in results between our study and those involving more experienced observers was at least partly attributable to the extensive training and detailed

TABLE 7. Mean CVs across all observers and species for six studies.

Studies						
This study[1]						
Markwood	Teakettle	Enemar (1962)[2]	Snow (1965)[3]	Hogstad (1966)[2]	Enemar et al. (1978)[2]	O'Connor and Marchant (1981)[1]
33%	37%	31%	34%	17%	35%	29%
			23%		29%	
			17%			
			28%			

[1] Based on number of territories for species that all observers detected.
[2] Based on number of detections for species that all observers detected.
[3] Based on number of territories for the 11 most common species.

TABLE 8. Mean CVs of territory totals (numbers pooled for all species) or contact totals for six studies.

This study[1]			Snow (1965)				
Markwood	Teakettle	Enemar (1962)	All species	Eleven most common species	Hogstad (1966)	Enemar et al. (1978)	O'Connor and Marchant (1981)[2]
6%	38%	15%	19%	29%	6%	13%	9%
			3%	4%		12%	
			12%	6%			
			23%	23%			

[1] Territory totals based on the mean of seven analysts.
[2] Territory totals based on the mean of three analysts.

instructions given to our observers before mapping began.

IMPLICATIONS FOR SAMPLE EFFORT

Most previous workers concluded that the magnitude of analyst and observer effects on density estimates from spot mapping are within acceptable limits for applications of the method. However, most of those analyses were based on pooled counts of all species, a procedure that masks species-specific variability. Such pooled data usually have lower mean CVs than those for individual species (e.g., 10 of 11 pairwise comparisons between corresponding values in Tables 7 and 8). Furthermore, most applications of spot-mapping data are species specific, involving trend analyses, comparisons between habitats or seasons, biomass computations, estimates of species diversity, and so on. Consequently, we believe that a complete assessment of analyst and observer bias on results of spot mapping should address individual species.

At the species level, CVs even of common species were generally high, although they were not exceptional for field data in ecology. Implications of these values must be considered by practitioners when applying results of spot mapping. Based on CVs, Svensson (1974; Table 8) showed that large numbers of plots must be sampled to detect levels of difference likely to be of interest to avian ecologists. Although Svensson's table was based on the generally accepted alpha level of 0.05, the power level used was only 0.50. Following his guidelines, one would commit a Type II error (i.e., fail to detect a real difference at the specified level) 50% of the time. Using the formula

$$d^2 = 2[z(1 - \text{alpha}) - z(1 - \text{power})]^2 \times CV^2/n$$

in which

d = the percentage difference detectable (to be conservative, divide the difference between two samples by the smaller of the two)
CV = coefficient of variation
n = the number of plots needed in each sample

TABLE 9. Number of plots needed in each sample to detect various levels of difference between samples, at various alpha and power levels, and assuming CV = 28%.

	Power											
	10% difference				20% difference				30% difference			
Alpha	0.95	0.90	0.85	0.80	0.95	0.90	0.85	0.80	0.95	0.90	0.85	0.80
					Two-tailed test							
0.05	204	165	141	123	51	41	35	31	23	18	16	14
0.10	170	134	113	97	42	34	28	24	19	15	13	11
0.15	149	116	96	82	37	29	24	20	17	13	11	9
0.20	134	103	84	71	34	26	21	18	15	11	9	8
					One-tailed test							
0.05	170	134	113	97	42	34	28	24	19	15	13	11
0.10	134	103	84	71	34	26	21	18	15	11	9	8
0.15	113	84	67	55	28	21	17	14	13	9	7	6
0.20	97	71	55	44	24	18	14	11	11	8	6	5

alpha = 0.05 (Type I error 5% of the time)
power = 0.80 (Type II error 20% of the time),
z(x) = value of a standard normal random variable that cuts off 100x% of the left tail (see Steel and Torrie 1980, p. 51–54, 578),

we recomputed all values in Svensson's table using a power level of 0.80 (i.e., committing a Type II error only 20% of the time). Estimated sample sizes in Svensson's table averaged only 59% (range = 54%–75%) of those needed for a power of 0.80. The mean of CVs shown in our Table 7 (species-specific summaries from six studies) is 28%, and these values came from only the more common species in each study. Using an alpha level of 0.05 and a power level of 0.80, one would need 31 plots in each sample to detect a significant difference of 20% between density estimates of a species with CV = 28%; to do so at a power level of 0.90 would require 41 plots in each sample. Sample sizes needed for other combinations of alpha and power levels (Table 9) indicate that spot mapping even for relatively common species is not a particularly useful method without at least 30 plots in each sample used in a comparison. And even this would not be sensitive to differences in density of less than 20%, unless one is willing to settle for an alpha level exceeding 0.05 and a power level of only 0.80. The situation is less demanding for one-tailed tests, but still one would need at least 20 plots in each sample (Table 9). O. Järvinen (pers. comm.) suggested that spot mapping may be insensitive to differences smaller than 100%.

These results clearly show the need for researchers to consider the variance of density estimates from spot mapping and the effect of this uncertainty on their power to detect statistical differences among plots. For example, a CV of 28% used to estimate needed sample sizes in Table 9 was probably optimistic. The mean CV for all species is probably closer to 50%, judging from results of O'Connor and Marchant (1981) and this study. CVs of 50% or larger occurred for 27% of the species at Markwood, 37% of the species at Teakettle, and 31% of the species reported by O'Connor and Marchant (1981). To detect a 10% difference (alpha = 0.05, power = 0.80) in density between two samples of such a species (CV = 50%) would require 393 plots in each sample. On the other hand, only 16 plots would be needed in each sample for a similar comparison of a species with CV = 10%, but even that would require an effort probably unavailable to most individual researchers. Furthermore, CVs ≤ 10% occurred for only 20% of the species at Markwood, 14% of the species at Teakettle, and 11% of the species reported by O'Connor and Marchant (1981).

Although sample-size requirements indicated by this study may be prohibitive, the situation is not hopeless. Our study design controlled effects of weather, day, season, year, and habitat on density estimates. Consequently, variability reported here was presumably associated only with observers and analysts. In studies of paired control and test plots in the same habitat type, researchers can control these sources of variability by using the same observer(s) on all plots to be compared, and balancing the effort of each observer over all plots. For example, we normally use 12 visits when spot mapping a plot. When comparing two or more plots, we have three observers complete four visits per plot or four observers complete three visits per plot. Results of each observer are pooled for a given plot, observers independently analyze the species maps, and any differences in density estimates among observers are discussed and resolved according to our rules for interpreting species maps. This gives a single consensus density estimate for each species and each plot, and it balances input from observers and analysts. Similar designs have been used to compare densities between plots by B. Noon (pers. comm.) in the Adirondack Mountains of New York and by R. J. Fuller (pers. comm.) in England. These measures, of course, add to the cost of what is already regarded as a very expensive method to estimate

TABLE 9. Extended.

Power							
40% difference				50% difference			
0.95	0.90	0.85	0.80	0.95	0.90	0.85	0.80
Two-tailed test							
13	10	9	8	8	7	6	5
11	8	7	6	7	5	5	4
9	7	6	5	6	5	4	3
8	6	5	4	5	4	3	3
One-tailed test							
11	8	7	6	7	5	5	4
8	6	5	4	5	4	3	3
7	5	4	3	5	3	3	2
6	4	3	3	4	3	2	2

bird densities. But results of this and similar studies indicate that such measures are needed.

Implications of observer and analyst variability need to be considered when using spot-mapping data for various purposes. The highly significant correlations between rank orders of species' relative abundance between observers indicate that spot mapping is suitable for such ranking purposes. Similarly, low CVs for the number of territories detected on the plots indicate that relatively few replicate plots are needed to compare bird species richness between habitats (e.g., see James and Wamer 1982). Use of spot mapping to monitor trends in bird populations, as currently done in England and Sweden, is an effective and valid procedure. Large numbers of plots are sampled in both countries; the same observers are involved year after year (plots where observers change between years are excluded from trend analyses); and both countries have procedures to reduce analyst variability (R. J. Fuller and S. Svensson, pers. comm.). Moreover, the number of plots needed to establish confidence in trends over time is less than needed to compare densities between different habitats at the same time, because the same plots are used for comparison from year to year. Finally, studies comparing densities of species between habitats, based on annual Breeding Bird Censuses in the U.S.A. as reported in *Audubon Field Notes* and *American Birds,* should be viewed with skepticism. The literature includes publications (e.g., MacArthur and MacArthur 1961) based on these results that have taken data from plots of different size, surveyed and analyzed by different observers, and done in different years. The extent to which these factors have obscured real effects is unknown, but we suspect that it may be substantial.

ACKNOWLEDGMENTS

C. C. Cutler, J. C. Lovio, B. R. North, R. J. Reynolds, J. L. Seay, and R. S. Suydam assisted with the fieldwork. J. A. Baldwin provided statistical consultation. D. G. Dawson, R. J. Fuller, O. Järvinen, B. R. Noon, and R. J. O'Conner offered useful, critical comments on the manuscript. To all we express our sincere appreciation.

LITERATURE CITED

ANONYMOUS. 1970. Recommendations for an international standard for a mapping method in bird census work, p. 49–52. *In* S. Svensson [ed.], Bird census work and environmental monitoring. Bull. Ecol. Res. Comm. 9, Swedish National Scientific Research Council, Stockholm.

BELL, B. D., C. K. CATCHPOLE, AND K. J. CORBETT. 1968. Problems of censusing Reed Buntings, Sedge Warblers and Reed Warblers. Bird Study 15:16–21.

BELL, B. D., C. K. CATCHPOLE, K. J. CORBETT, AND R. J. HORNBY. 1973. The relationship between census results and breeding populations of some marshland passerines. Bird Study 20:127–140.

BEST, L. B. 1975. Interpretational errors in the "mapping method" as a census technique. Auk 92:452–460.

DESANTE, D. F. 1981. A field test of the variable circular-plot censusing technique in a California coastal scrub breeding bird community. Stud. Avian Biol. 6:177–185.

DESANTE, D. F. 1986. A field test of the variable circular-plot censusing method in a Sierran subalpine forest habitat. Condor 88:129–142.

DICKSON, J. G. 1978. Comparison of breeding bird census techniques. Am. Birds 32:10–13.

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

EMLEN, J. T. 1977. Estimating breeding bird densities from transect counts. Auk 94:455–468.

ENEMAR, A. 1959. On determination of the size and composition of a passerine bird population during the breeding season. Vår Fågelvärld, Suppl. 2:1–114.

ENEMAR, A. 1962. A comparison between the bird census results of different ornithologists. Vår Fågelvärld 21:109–120.

ENEMAR, A., B. SJÖSTRAND, AND S. SVENSSON. 1978. The effect of observer variability on bird census results obtained by a territory mapping technique. Ornis Scand. 9:31–39.

FRANZREB, K. E. 1976. Comparison of variable strip transect and spot-map methods for censusing avian populations in a mixed-coniferous forest. Condor 78:260–262.

FRANZREB, K. E. 1981. A comparative analysis of territorial mapping and variable-strip transect censusing methods. Stud. Avian Biol. 6:164–169.

HAMEL, P. B. 1984. Comparison of variable circular-plot and spot-map censusing methods in temperate deciduous forest. Ornis Scand. 15:266–274.

HILDÉN, O. 1981. Sources of error involved in the Finnish line transect method. Stud. Avian Biol. 6:152–159.

HOGSTAD, O. 1966. Factors influencing the efficiency of the mapping method in determining breeding bird populations in conifer forests. Nytt. Mag. Zool. (Oslo) 14:125–141.

HOWELL, J. C. 1951. The roadside census as a method of measuring bird populations. Auk 68:334–357.

JAMES, F.C., AND N. O. WAMER. 1982. Relationships between temperate forest bird communities and vegetation structure. Ecology 63:159–171.

JÄRVINEN, O., R. A. VÄISÄNEN, AND A. ENEMAR. 1978a. Efficiency of the line transect method in mountain birch forest. Ornis Fenn. 55:16–23.

Järvinen, O., R. A. Väisänen, and W. Walankiewicz. 1978b. Efficiency of the line transect method in central European forests. Ardea 66:103–111.

MacArthur, R. H., and J. W. MacArthur. 1961. On bird species diversity. Ecology 42:595–598.

O'Connor, R. J. 1981. The influence of observer and analyst efficiency in mapping method censuses. Stud. Avian Biol. 6:372–376.

O'Connor, R. J., and J. H. Marchant. 1981. A field validation of some Common Birds Census techniques. Report from the British Trust for Ornithology to the Nature Conservancy Council, Contract No. HF3/03/192.

O'Meara, T. E. 1981. A field test of two density estimators for transect data. Stud. Avian Biol. 6: 193–196.

SAS Institute. 1982. SAS user's guide: statistics. SAS Institute, Cary, NC.

Snow, D. W. 1965. The relationship between census results and the breeding population of birds on farmland. Bird Study 12:287–304.

Steel, R.G.D., and J. H. Torrie. 1980. Principles and procedures of statistics. McGraw-Hill Book Co., New York.

Stewart, R. E., J. B. Cope, C. S. Robbins, and J. W. Brainerd. 1952. Seasonal distribution of bird populations at the Patuxent Research Refuge. Am. Midl. Nat. 47:257–263.

Svensson, S. 1974. Interpersonal variation in species map evaluation in bird census work with the mapping method. Acta Ornithol. 14:322–338.

Verner, J., and L. V. Ritter. 1988. Comparison of transects and spot mapping in oak-pine woodlands of California. Condor 90:401–419.