Ann. Zool. Fennici 26:191–199. 1989

# Coping with sources of variability when monitoring population trends

Jared Verner & Kathleen A. Milne

Verner, J. & Milne, K. A. 1989: Coping with sources of variability when monitoring population trends. — Ann. Zool. Fennici 26:191–199.

Seven observers independently sampled the bird assemblage in oak-pine woodlands in central California by completing 5-min point counts at 210 counting stations during April 1986. Either total counts or frequencies, or both, could be used for monitoring trends in relative abundance of birds. However, density estimates have limited usefulness for most monitoring efforts. Location and date affected results significantly, even in relatively homogeneous habitat and with all counts being completed in a 9-day period. We see little opportunity to control these sources of variation more than was accomplished in this study. Results showed that the observer was a major source of variation in counts and frequencies. We recommend (1) that three or more observers sample all sites in a system to monitor trends in bird populations on the small geographic scale represented by this study, (2) that observers pass a performance test, including hearing, to be included on a monitoring field team, and (3) that all observers undergo training prior to field sampling. However, even these measures will not eliminate all effects of observer variability, requiring caution by practitioners when interpreting results.

*Jared Verner and Kathleen A. Milne, Pacific Southwest Forest and Range Experiment Station, USDA Forest Service, 2081 East Sierra Avenue, Fresno, California 93710, USA.*

## 1. Introduction

Several laws recently enacted in the United States require certain agencies to monitor trends in various renewable resources (Salwasser et al. 1983). As a result, much interest has developed concerning cost-effective methods that deliver reliable estimates of the relative abundance of specified resources (e.g., Verner 1983). At our laboratory in Fresno, California, we have been studying sources of bias in the three methods used most commonly to estimate the abundance of birds — spot mapping, transects, and point counts. Because we believe these methods cannot deliver reasonably accurate density estimates of most species in most habitats (e.g., see Verner 1985, Verner & Ritter 1985), we have focused our attention on total counts and frequencies as potential measures of trends in bird populations. Our methods are based on modifications of the I.P.A. and E.F.P. methods (Blondel et al. 1970, 1981; Ferry 1974; Blondel 1975, 1977) developed and used extensively by workers in France and elsewhere.

This paper reports results from one complete sample of a test system of point counts installed in oak-pine woodlands at the San Joaquin Experimental Range, Madera County, in central California, to monitor yearly variations in bird populations. The system is applicable to areas of medium to large size (>1800 ha); smaller areas would not accommodate the number of counting stations needed at a minimum distance of 300 m between stations.

Primary objectives of this study were (1) to measure the contribution of observers, dates, and locations to total variability in counts and frequencies of birds in oak-pine woodlands, and (2) to compare total counts and frequencies as measures for detecting trends.

## 2. Study area

The San Joaquin Experimental Range (SJER) is an area of 1875 ha, ranging in elevation from 215 to 520 m, in the western foothills of the Sierra Nevada of California. The climate is characterized by cool, wet winters and hot, dry summers. Annual precipitation averages 48.6 cm (43-yr mean, 1935–77), most falling as rain from November through March. A sparse woodland overstory of blue oak (*Quercus douglasii*), digger pine (*Pinus sabiniana*), and interior live oak (*Q. wislizenii*) covers most of SJER. An understory of scattered shrubs includes mainly buck brush (*Ceanothus cuneatus*), chapparral whitethorn (*C. leucodermis*), redberry (*Rhamnus crocea*), and Mariposa manzanita (*Arctostaphylos mariposa*). In a few smaller patches, the overstory is primarily blue oak, and a shrub understory is meagre or missing. Some areas of typical annual grasslands extend throughout the remainder of SJER where the overstory and understory are not dense enough to shade them out or are lacking altogether.

## 3. Methods

Seven lines with 30 counting stations each were established primarily in oak-pine habitats throughout SJER, with the aid of aerial photos and topographic maps (scale = 13 500:1; contour interval = 38 m). Counting stations were at least 200 m apart along the same line and between the separate lines. (Closer spacing than is ideal for independent samples was used here to allow six counts per hour.) All counting stations were clearly identified by placement of large plastic tags wired to fences, trees, shrubs, and occasionally to steel fence posts set in open areas specifically for that purpose. Numerous additional tags placed between stations along a line gave directions for continuing along the line; a "tour guide" described in detail the location of each tag, what it was wired to, and the distance and direction to the next tag along the line. With this system, observers unfamiliar with the lines were able to follow them quickly and accurately.

Recording of birds along a line began at the first station on the line at 10 min after official sunrise. The counting period was 5 min, after which the observer moved quickly to the next station and began counting at exactly 10 min after counting began at the first station. By adhering to this schedule, an observer recorded birds at 6 stations per hour, so all 30 stations on a line were sampled within 5 hr, and the entire system was sampled by each observer in seven mornings. Counts were not done during rainy mornings, and counts done during days when wind consistently exceeded 32 km/h (by Beaufort scale) were repeated the following count day. Windy-day counts were not included in the present analysis.

Results reported here were taken from 7–15 April 1986, by seven observers randomly assigned to lines (lattice design) in such a way that none sampled the same line on the same day and all sampled all seven lines. Observers were carefully selected to be expert birders and especially to be expert in the identification of birds at SJER by sight and sound. At each station, they recorded the date, time, wind velocity, percent cloud cover, rain activity, and temperature. For each bird detected, they recorded the species, cue (visual, song, call, or other sound) first detected, distance (in meters) from the counting point, age (adult or juvenile), and whether the bird had probably been detected from a previous point that morning. Each observer's hearing was tested within a week after the field work was completed.

In addition to the seven primary observers, two other observers began counting 20 minutes after the primary observer on the line to which they were randomly assigned each day. Extensive analysis of the results from these observers revealed differences no greater than those observed among the seven primary observers. Consequently, their results were included in an analysis of the gain in precision of counts and frequencies to be expected when pooling data from increasing numbers of observers. Their data were not used in other analyses reported here, however, because those analyses required a design balanced with respect to the number of observers, lines, and days.

Based on extensive knowledge of birds at SJER and in oak-pine woodlands of this portion of the western Sierra Nevada in general, bird species were grouped into the following four categories for analysis:

1) *breeders* have been found breeding at SJER either during the year of this study or during previous years;
2) *winter residents* have regularly spent 2 or more months at SJER during the winter but do not breed there;
3) *transients* have regularly migrated through SJER in spring, fall, or both, but have not remained for extended periods;
4) *casuals/aquatics* were
   a) known to breed regularly or spend the winter months in areas near enough to SJER that their daily movements occasionally included SJER, where they were detected infrequently, and/or
   b) associated with aquatic habitats and therefore not necessarily dependent upon the oak-pine woodlands that were the target of the monitoring effort.

The analysis addresses sources of variation in counts, frequencies, and species richness attributable to observer, line (pooling results from the 30 counting stations per line for each observer each day), and date. Because lines were in different parts of the oak-pine woodland, line effects are referred to hereafter as "location" effects. Statistical tests are identified in the results section, as appropriate; statistical significance has been arbitrarily set at a probability level of 0.05.

## 4. Results

### 4.1. Counts

The seven observers tallied a total of 24 411 birds of 97 species. Breeders comprised the largest proportion of these — 85.9% (Table 1); 49 of 52 species known to breed at SJER were detected. Species not detected were nocturnal or nested uncommonly around buildings in the headquarters area. Winter residents made up an additional 11.6% of the total count, and 14 of 18 species known to spend the winter on a regular basis were detected. Transients (19 of 35 known transient species detected) contributed 2.0% of the total count, and casuals/aquatics (15 of 71 species detected) made up the remaining 0.5%. Because casuals/aquatics were not a primary target of this monitoring effort, and because they comprised such a small proportion of the total sample, they have been excluded from further analysis here.

Table 1. Total counts of breeders, winter residents, and transients, by observer and date (April).

| Date: | 7 | 8 | 9 | 10 | 11 | 14 | 15 | Total | Mean ± 2*SE* | %*CV* |
|---|---|---|---|---|---|---|---|---|---|---|
| Breeders | | | | | | | | | | |
| Observer 1 | 518 | 455 | 468 | 555 | 502 | 465 | 440 | 3403 | 486.1 ± 30.1 | 8.3 |
| 2 | 438 | 431 | 475 | 462 | 490 | 419 | 416 | 3131 | 447.3 ± 21.3 | 6.4 |
| 3 | 452 | 511 | 417 | 466 | 461 | 515 | 446 | 3268 | 466.9 ± 26.1 | 7.5 |
| 4 | 246 | 284 | 279 | 292 | 269 | 252 | 301 | 1923 | 274.7 ± 15.0 | 7.4 |
| 5 | 431 | 407 | 430 | 390 | 434 | 371 | 455 | 2918 | 416.9 ± 21.5 | 7.0 |
| 6 | 362 | 395 | 403 | 408 | 478 | 381 | 408 | 2835 | 405.0 ± 26.8 | 8.9 |
| 7 | 449 | 457 | 476 | 507 | 544 | 502 | 551 | 3486 | 498.0 ± 29.6 | 8.0 |
| Total | 2896 | 2940 | 2948 | 3080 | 3178 | 2905 | 3017 | 20964 | 2994.9 ± 76.7 | 3.5 |
| Winter residents | | | | | | | | | | |
| Observer 1 | 100 | 73 | 44 | 51 | 83 | 62 | 35 | 448 | 64.0 ± 17.0 | 35.8 |
| 2 | 57 | 63 | 38 | 16 | 46 | 79 | 73 | 372 | 53.1 ± 16.1 | 40.9 |
| 3 | 23 | 37 | 57 | 107 | 117 | 30 | 46 | 417 | 59.6 ± 27.8 | 63.0 |
| 4 | 27 | 26 | 43 | 56 | 46 | 43 | 18 | 259 | 37.0 ± 10.0 | 36.5 |
| 5 | 79 | 58 | 38 | 68 | 72 | 35 | 17 | 367 | 52.4 ± 16.9 | 43.6 |
| 6 | 105 | 64 | 93 | 70 | 57 | 48 | 52 | 489 | 69.9 ± 15.9 | 30.7 |
| 7 | 65 | 82 | 69 | 109 | 58 | 50 | 55 | 488 | 69.7 ± 15.0 | 29.0 |
| Total | 456 | 403 | 382 | 477 | 479 | 347 | 296 | 2840 | 405.7 ± 51.5 | 17.1 |
| Transients | | | | | | | | | | |
| Observer 1 | 8 | 9 | 16 | 6 | 12 | 18 | 11 | 80 | 11.4 ± 3.2 | 37.8 |
| 2 | 5 | 5 | 5 | 11 | 15 | 20 | 16 | 77 | 11.0 ± 4.6 | 56.3 |
| 3 | 12 | 10 | 8 | 9 | 10 | 15 | 14 | 78 | 11.1 ± 1.9 | 23.4 |
| 4 | 4 | 5 | 8 | 10 | 15 | 12 | 12 | 66 | 9.4 ± 3.0 | 42.4 |
| 5 | 9 | 8 | 14 | 19 | 14 | 24 | 7 | 95 | 13.6 ± 4.6 | 45.9 |
| 6 | 4 | 7 | 1 | 5 | 16 | 13 | 1 | 47 | 6.7 ± 4.3 | 86.3 |
| 7 | 5 | 5 | 4 | 5 | 9 | 13 | 12 | 53 | 7.6 ± 2.8 | 49.3 |
| Total | 47 | 49 | 56 | 65 | 91 | 115 | 73 | 496 | 70.9 ± 18.3 | 34.9 |

Table 2. Total counts of a selected subset of breeding species to illustrate variability in counts by different observers.

| Observer: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | %CV |
|---|---|---|---|---|---|---|---|---|
| Turkey vulture (*Cathartes aura*) | 56 | 107 | 57 | 41 | 59 | 66 | 25 | 43.1 |
| American kestrel (*Falco sparverius*) | 6 | 11 | 7 | 9 | 0 | 2 | 3 | 72.8 |
| Mourning dove (*Zenaida macroura*) | 165 | 165 | 128 | 120 | 76 | 143 | 96 | 26.3 |
| Greater roadrunner (*Geococcyx californianus*) | 4 | 0 | 7 | 2 | 0 | 12 | 8 | 95.4 |
| Black phoebe (*Sayornis nigricans*) | 3 | 2 | 2 | 1 | 2 | 1 | 4 | 49.9 |
| Scrub jay (*Aphelocoma coerulescens*) | 163 | 150 | 183 | 195 | 167 | 182 | 122 | 14.7 |
| Plain titmouse (*Parus inornatus*) | 271 | 262 | 247 | 424 | 338 | 288 | 148 | 30.0 |
| Bushtit (*Psaltriparus minimus*) | 119 | 296 | 45 | 99 | 97 | 53 | 66 | 77.7 |
| White-breasted nuthatch (*Sitta carolinensis*) | 130 | 107 | 137 | 183 | 105 | 152 | 97 | 23.4 |
| Rock wren (*Salpinctes obsoletus*) | 19 | 8 | 10 | 11 | 3 | 16 | 9 | 48.6 |
| Bewick's wren (*Thryomanes bewickii*) | 141 | 106 | 144 | 187 | 22 | 117 | 105 | 43.3 |
| House wren (*Troglodytes aedon*) | 133 | 133 | 147 | 136 | 130 | 126 | 96 | 12.3 |
| Western bluebird (*Sialia mexicana*) | 80 | 117 | 77 | 76 | 37 | 50 | 31 | 44.6 |
| Brown towhee (*Pipilo fuscus*) | 110 | 82 | 75 | 55 | 86 | 75 | 63 | 22.7 |
| Brewer's blackbird (*Euphagus cyanocephalus*) | 0 | 41 | 13 | 46 | 5 | 13 | 4 | 106.0 |
| Lesser goldfinch (*Carduelis psaltaria*) | 157 | 378 | 165 | 207 | 186 | 136 | 75 | 50.6 |

Total counts of various bird species were highly variable for the different observers (Table 2). Among breeders, for example, only 36 of 49 species were recorded by all seven observers, three were recorded by six observers, two by five, two by four, one by three, three by two, and two by one observer. Among the 36 species recorded by all observers, the lowest count of a given species by any observer averaged

Table 3. Mean percent coefficient of variability of counts attributable to date + location, observer + location, and observer + date for three categories of bird.

| | Breeders | Winter residents | Transients |
|---|---|---|---|
| Date + location | 7.6 | 39.9 | 48.8 |
| Observer + location | 18.9 | 41.1 | 41.6 |
| Observer + date | 19.0 | 33.3 | 50.8 |

only 35.8% of the highest count ($SD$ = 16.6; range = 9.1% to 72.7%); the lowest count exceeded 50% of the highest for only 7 of the 36 species.

Precision in total daily counts, as measured by the coefficient of variation ($CV$), was greatest for breeders, followed by winter residents, and transients. The mean $CV$ for the 49 breeding species was 66.5%; it was 171.4% for the 17 winter residents, and 171.6% for the 17 transients. This trend was even more clearly shown when results were pooled by location (Table 3). The combined effects of observer and location (date effects held constant) and of observer and date (location effects held constant) each gave mean $CV$s 2.5 times that of the combined effects of date and location (observer effects held constant) (Table 3), suggesting that observers contributed most to variability in counts of breeders. In fact, a mean $CV$ of 7.6% for the combined effects of date and location suggest sampling from acceptably homogeneous populations. On the other hand, location effects apparently contributed most to variability in counts of winter residents, and date effects were most important for transients (Table 3). These relationships were confirmed by a three-way analysis of variance (Table 4). Only observer effects had a significant $F$ value for breeders, and 11 of 21 pairwise comparisons of observers were significantly different. Although all three variables had significant $F$ values for winter residents, that for location effects was greatest. For transients, both observers and dates had significant $F$ values, but date effects were greater.

Assuming independent sets of observers, the reduction in $CV$ of total counts achievable by pooling data from two or more observers was estimated using the formula

$$CV = \frac{\sqrt{Var/n}}{\bar{x}}$$

where $Var$ = the variance of a given sample, $n$ = the number of observers to be used, and $\bar{x}$ = the sample mean.

Table 4. Three-way analysis of variance of observer, location, and date effects on total counts.

| | $df$ | $F$ | $P>F$ | $R^2$ | Differences[1] |
|---|---|---|---|---|---|
| Breeders | | | | | |
| Overall | 18 | 14.23 | 0.0001 | 0.895 | |
| Observer | 6 | 39.84 | 0.0001 | | 11 |
| Location | 6 | 1.31 | 0.2822 | | 0 |
| Date | 6 | 1.52 | 0.2041 | | 0 |
| Winter residents | | | | | |
| Overall | 18 | 6.51 | 0.0001 | 0.796 | |
| Observer | 6 | 4.71 | 0.0017 | | 3 |
| Location | 6 | 11.38 | 0.0001 | | 9 |
| Date | 6 | 3.44 | 0.0105 | | 2 |
| Transients | | | | | |
| Overall | 18 | 3.50 | 0.0012 | 0.677 | |
| Observer | 6 | 2.99 | 0.0208 | | 1 |
| Location | 6 | 0.95 | 0.4764 | | 0 |
| Date | 6 | 6.56 | 0.0002 | | 5 |

[1]Number of significant pairwise differences, Tukey's studentized range test.

The mean $CV$ decreases about 29% when the number of observers is increased from one to two and another 13% when data from three observers are pooled, but the rate of gain in precision declines steadily with further addition of observers (Fig. 1, curves A and D).

Using pooled results from all possible combinations of three observers (therefore, not independent subsets of observers), the $CV$ of total counts declined rapidly as counts increased to a total of about 70, but only slowly with increasing counts beyond that (Fig. 2). For example, the mean $CV$ for all breeders was 31.2% ($n$ = 49); for species with counts exceeding 70 it was only 16.9% ($SD$ = 7.45; range = 4.5% to 31.8%; $n$ = 26), but it was 47.4% ($SD$ = 34.18; range = 11.2% to 142.7%; $n$ = 23) for species with counts of 70 or less.

#### 4.2. Frequencies

Frequency (number of stations at which a species was detected, divided by the total number of stations sampled) was examined as an alternative to total count to indicate a species' relative abundance. The rank order of relative abundance of species as measured by frequencies was significantly correlated to that of total counts (Spearman's $\rho > 0.99$, $P < 0.001$, for all observers separately). Although not as variable

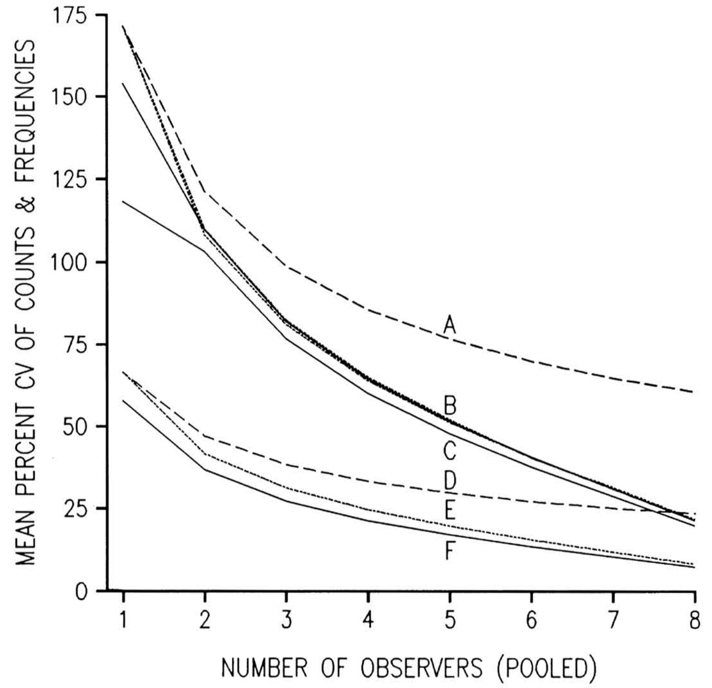


Fig. 1. Mean percent *CV* of total counts and frequencies of species by category, as a function of the number of observers whose data were pooled. Dashed lines give results for counts assuming that all subsets of observers are independent (using the formula in section 4.1); dotted lines give results for counts using all possible subsets of observers in this study (not all independent subsets); and solid lines give results for frequencies using all possible subsets of observers in this study. — A = winter residents and transients (identical results); B = overlapping results for winter residents and transients (dotted lines) and transients (solid line); C = winter residents; D, E, and F = breeding species.

as total counts, frequencies were still surprisingly variable for the different observers, especially considering the fact that frequencies required only presence/absence data. The difference between the lowest and highest frequencies among the seven observers averaged 45.9% (*SD* = 20.3; range = 9.6% to 84.0%) for the 36 species detected by all observers, and the lowest frequency exceeded 50% of the highest in only 14 of those cases. The precision of frequencies was greatest for breeders (mean *CV* = 57.8%), followed by winter residents (mean *CV* = 118.13%), and transients (mean *CV* = 153.9%).

The reduction in mean *CV* of frequencies achievable by increasing the number of observers was estimated by pooling data for all possible combinations of two (*n* = 36), three (*n* = 84), four (*n* = 126), five (*n* = 126), six (*n* = 84), seven (*n* = 36), and eight (*n* = 9) observers (Fig. 1). However, because the various subsets of observers were not independent, the apparent reduction in *CV* with increasing numbers of observers was greater than one should expect if all subsets had different observers. To compare the reduction in *CV*s between frequency and count data with increasing numbers of observers, therefore, we also calculated the mean *CV*s of count data for all possible subsets of two, three, four, five, six, seven, and eight observers. Reduction in the *CV* of count data by this method was only slightly less than that for frequency data (Fig. 1). The difference between the *CV*s for counts using the formula (Fig. 1, curves A and D) and those empirically derived from all possible combinations of subsets of the observers (Fig. 1, curves B and E) may approximate the bias introduced by not using independent subsets of observers.

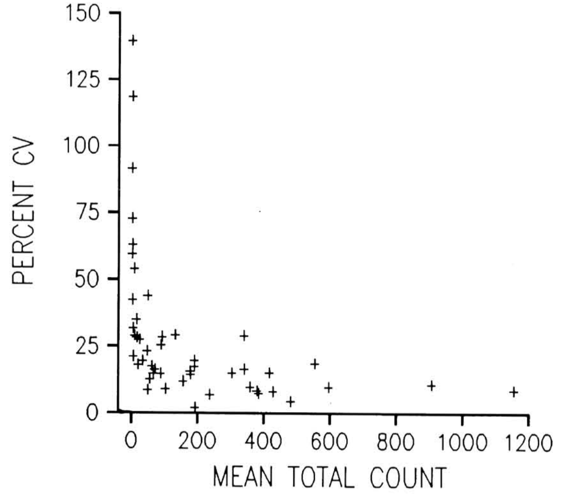


Fig. 2. Percent *CV* of total counts of all possible subsets of three observers, in relation to mean total count.

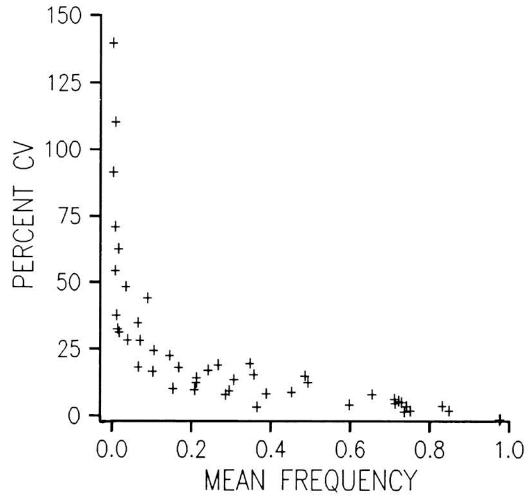


Fig. 3. Percent *CV* of frequencies of all possible subsets of three observers, in relation to mean frequency.

Using pooled results from all possible combinations of three observers, the *CV* of frequencies declined rapidly as frequency increased from 0.01 to 0.10, but only slowly with increasing frequency beyond 0.10 (Fig. 3). For example, the mean *CV* for all breeders was 27.2% ($n$ = 49); for species with frequencies exceeding 0.10 it was only 12.2% ($SD$ = 6.88; range = 1.1% to 26.9%; $n$ = 33), but it was 58.0% ($SD$ = 33.30; range = 20.7% to 142.7%; $n$ = 16) for species with frequencies of 0.10 or less.

Although the same or nearly the same bird assemblage was sampled at the same sites during the same 7 days, the combined variation attributable to observers and dates was enough that in 16.6% of all possible pairwise comparisons of breeding species by single observers, the observed frequencies of a given species were significantly different (untransformed frequency values, using Tukey's HSD test, which adjusts for multiple comparisons). In the extreme case of the California quail (*Callipepla californica*), 66.7% of all pairwise comparisons were significantly different. In addition, the frequencies of several species were too low to permit detection of significant differences between observers (i.e., insufficient power). Significant differences were found in more than 10% of the pairwise comparisons of single observers for 25 of the 49 breeding species. Finally, frequency values and the percent of significant differences in all pairwise comparisons between single observers were significantly correlated ($r = 0.53$; $P < 0.001$; $n = 49$), indicating that we more often erred by detecting significant differences between frequencies of common than of uncommon species. This relationship was apparently driven by the fact that we lacked power to detect differences at low frequencies, as the correlation coefficient (–0.04) for frequencies of 0.15 and greater was not statistically significant.

Although pooling data among all possible subsets of three observers markedly reduced the number of significant differences found between pairs of frequency values, it did not eliminate all errors of this type. For example, only 4.7% of all comparisons between pairs of pooled data sets of three observers were significantly different among the breeding species, and significant differences occurred in more than 10% of the pairwise comparisons for only 9 of the 49 breeders. However, in the extreme case of the western bluebird, 34.3% of all pairwise comparisons were significantly different. The correlation between frequency values and the percent of significant differences in all pairwise comparisons was not significant ($r = 0.19$; $P > 0.05$; $n = 49$). We do not know the extent to which the lack of independence among the pooled data sets of all combinations of three observers biased our assessment of the benefits of pooling, but undoubtedly those benefits are not as great as appears in the present analysis.

## 5. Discussion

Separate analysis of results for breeders, winter residents, and transients was a useful procedure that gave insights into point counting as a method for monitoring population trends that we would have missed had we analyzed all species together, irrespective of category. For example, we were encouraged that breeders were counted with the greatest precision, because they are most likely to be emphasized in a long-term monitoring program. Breeders are typically territorial and males tend to be vocal, making them more likely to be detected in the same locality day after day. Not surprisingly, then, we found no significant location and date effects on counts of breeders. On the other hand, significant observer effects were common and often large.

The fact that counts of transients were most affected by date probably resulted from daily differences in the abundance and species composition of migrants passing through the area. Counts of winter residents were apparently most influenced by location effects. This probably resulted because winter residents tend to occur in flocks, or perhaps they actually are more selective in their choice of microhabitats than breeders or transients. These patterns, although amenable to reasonable biological interpretations, are based on a single replication of this study. The consistency of the patterns will be tested with additional replications using the same study design.

A monitoring system such as that described here, involving a large number of counting stations, gives both counts and frequencies as possible measures of trends. Both measures should be used, although frequencies will likely prove to be more sensitive measures of trends than total counts, in part because they can be measured more precisely than total counts.

In small-scale monitoring efforts, we would recommend against using density estimates from point counts or transects for monitoring trends in bird populations. First, because trends can be adequately indicated by measures of relative abundance, such as total counts or frequencies, density estimates go beyond what is needed (see Verner 1985:249–252, Raphael 1987). Second, estimating the density of a species from point counts or transects using line-transect algorithms requires a total count not normally attained for all species detected. Burnham et al. (1980) recommended a count of 40 or more individuals to estimate densities with line transect methods, but they suggested that 60 to 80 individuals would be preferrable. At another point in their monograph, they stated that "Even with sample sizes of 100, one has difficulty in inferring the true underlying detection function" (Burnham et al. 1980:177). A recent study at our laboratory indicated that a sample of at least 100 individuals is needed to estimate densities with line transect methods (Verner and Ritter 1988). And even with samples that large, one can probably be fairly confident only in the precision of density estimates, not in their accuracy. Thus, unless a monitoring system involves a sufficient effort to attain counts of 100 or more of all species detected, or if one is willing to monitor trends only of those species whose counts reach 100, density estimates have limited usefulness for monitoring trends in bird populations.

Alternatively, one could estimate densities from frequencies, which are correlated with densities (Blondel 1975:583). However, the relationship is less precise with higher than with lower frequencies, leading Blondel et al. (1981:415) to caution that "It is not possible to infer densities from the frequencies." Trends can certainly be monitored more accurately with total counts and frequencies than with density estimates.

Precision in measures used to detect trends in populations must be a primary objective of any well-designed monitoring system. We selected point counting as the preferred method in this study, because (1) it allowed us to exactly specify the effort spent counting birds (e.g., see Svensson 1977), and (2) it allowed us to obtain a large number of samples in a reasonably short period of time (e.g., see Blondel 1975, Reynolds et al. 1980, Blondel et al. 1981, Dawson 1981). The latter is especially important when frequency is used as a measure of trend, because the number of frequency intervals equals the number of sites sampled. Studies with few sites do not give a reasonably fine resolution of relative abundance by frequency. Note, also, that frequencies are "unsuitable for comparing studies differing in time spent sampling each site or differing in plot sizes" (Verner 1985). Thus, frequency cannot be used as a measure of relative abundance with transect or spot-mapping methods.

Results showed that all effects measured — observer, location, and date — influenced some measure of abundance. Although we attempted to locate all counting stations in similar habitat, this was not always possible. The large number of stations needed, and the spacing criteria used to locate them, resulted unavoidably in some variation in habitats among the stations. However, general patchiness in the structure and composition of nearly all habitats with which we are familiar in North America leads us to believe that one would not often attain less variability in the locations of counting stations than we achieved in this study. Similarly, little, if anything, could be done to reduce the effect of date on counts of birds. All counts in this study were done within a 9-day period. This was accomplished by each observer's completing counts at 30 stations per day, which was about the limit (1) in terms of the optimum counting period during the day and (2) in terms of an observer's tendency to become fatigued (but see Bart & Schoultz 1984).

Observer variability has been documented in many previous studies (e.g. Carney & Petrides 1957, Enemar 1962, Emlen 1971, Berthold 1976, Kepler & Scott 1981, Scott et al. 1981), but the results of Bart (1985) are especially interesting in the present case. In a careful study involving several "expert" birders who sampled the same known assemblage of singing birds, Bart concluded that under-counting, over-counting, and misidentification all contributed to observer differences. "The average number of birds missed per period (with 20 actually present) was 6.0 (range among surveyors = 4.2–10.8). The average number over-counted was 1.0 (range = 0.2–4.1). The average number mis-identified was 0.6 (range = 0.16–2.1)" (Bart 1985:163). Svensson (1977) also reported considerable observer variation in a test of point counts for monitoring trends in Swedish bird populations, although his study design did not allow separation of observer effects from those of habitats and days. He proposed a calibration test of the observers prior to initiation of field work, with results of the test being used to deny participation in counts by observers who perform poorly. The value of training observers prior to field work has been well documented by Kepler & Scott (1981) and Scott et al. (1981). Dawson (1981) recommended using several observers so their differences can measured and allowed for during analysis (also see Scott et al. 1986).

The most obvious solution to observer variability in a monitoring program is to use the same observer(s) throughout the life of the program. Such an option is feasible for large-scale monitoring efforts, because the subset of sites at which observers change between years can be excluded from trend estimates between those years. This approach is used, for example, in the national-level monitoring programs in England (R. J. Fuller, pers. comm.) and Sweden (S. Svensson, pers. comm.). The annual Breeding Bird Survey in the United States includes observers as a covariate in the analysis of trends (C. S. Robbins, pers. comm.). These are feasible approaches for dealing with inter-observer variability. However, even when the same observer participates for many successive years, biases resulting from (1) improving observer skill and (2) declining auditory acuity with observer age may require other sorts of adjustments (e.g., see figure 3 in Faanes & Bystrak 1981).

In relatively small-scale monitoring efforts like that studied here, so few observers are involved that excluding results when observers change between years may have too large an impact on sample size to be tolerated. Therefore, the design must allow for control and/or measurement of observer variability. To accomplish this, we recommend that all suggestions proposed by Svensson (1977), Dawson (1981) Kepler & Scott (1981), and Scott et al. (1986) be followed. Specifically, a well-designed monitoring program should include (1) prior testing of observers' skills, including hearing (e.g., Emlen & DeJong 1981, Ramsey & Scott 1981), (2) prior training, and (3) multiple observers, all of whom sample all counting stations. The number of observers used would no doubt be limited by available funds, so a compromise is needed between increasing the cost and decreasing the variation in measures of abundance by adding observers. Based on results of this study, we tentatively recommend three observers, because the gain in precision was relatively large compared to one or two observers, but further gain from adding more observers was probably not worth the added cost (e.g., see Fig. 1, curves A and D). Even these procedures, however, will not completely eliminate the effects of observer variability. Therefore, when interpreting results, practitioners monitoring trends in bird populations must be mindful of this and other sources of variation in bird counts.

Our results show that the magnitude of observer variability can often result in erroneous conclusions about yearly differences in populations of common species. These may not be serious errors, because common species are not likely to be in jeopardy. On the other hand, data obtained from a point-counting system similar to that described here will not often allow reliable detection of trends in populations of uncommon birds. In this study, results from species delivering mean frequencies of 0.10 or less, or total counts of 70 or less (mean of about 0.1 bird per counting station) in data sets pooled across three different observers were far too imprecise. Unfortunately, however, these uncommon species are the ones about which we are most likely to be concerned and for which we are most likely to need reliable estimates of population trends. Detection of these trends will require separate monitoring systems designed specifically for the bird species in question, as need arises.

*Acknowledgements.* Laura B. Arnow, Christopher R. Cutler, Zev Labinger, Peter B. Landres, John C. Lovio, Barbara R. North, Richard J. Reynolds, and Robert S. Suydam assisted with the field work. John R. Probst, Martin G. Raphael, David Rugg, and J. Michael Scott provided constructive comments on the draft manuscript. James A. Baldwin and Charles J. Evans provided statistical consultation and assistance with data analysis. Gary P. Eberlein produced the graphics. To all we express our sincere appreciation.

## References

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

Bart, J. & Schoultz, J. D. 1984: Reliability of singing bird surveys: Changes in observer efficiency with avian density. — Auk 101:307–318.

Berthold, P. 1976: Methoden der Bestandserfassung in der Ornithologie: Übersicht und kritische Betrachtung. — J. Ornithol. 117:1–69.

Blondel, J. 1975: L'analyse des peuplements d'oiseaux, élément d'un diagnostic écologique. 1. La méthode des échantillonnages fréquentiels progressifs (E.F.P.). — Terre et Vie 29:533–589.

—"— 1977: The diagnosis of bird communities by means of frequential sampling (E.F.P.). — Pol. Ecol. Stud. 3:19–26.

Blondel, J., Ferry, C. & Frochot, B. 1970: La méthode des indices punctuels d'abondance (IPA) ou des relevés d'avifaune par "stations d'écoute." — Alauda 38:55–71.

—"— 1981: Point counts with unlimited distance. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:414–420. Cooper Ornithological Society.

Burnham, K. P., Anderson, D. R. & Laake, J. L. 1980: Estimation of density from line transect sampling of biological populations. — Wildl. Monogr. 71:1–102.

Carney, S. M. & Petrides, G. A. 1957: Analysis of variation among participants in pheasant cock-crowing censuses. — J. Wildl. Managem. 21:392–397.

Dawson, D. G. 1981: Counting birds for a relative measure (index) of density. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:12–16. Cooper Ornithological Society.

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

Emlen, J. T. & DeJong, M. J. 1981: The application of song detection threshold distance to census operations. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:346–352. Cooper Ornithological Society.

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

Faanes, C. A. & Bystrak, D. 1981: The role of observer bias in the North American Breeding Bird Survey. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:353–359. Cooper Ornithological Society.

Ferry, C. 1974: Comparison between breeding bird communities in an oak forest and a beech forest, censused by the IPA method. — Acta Ornithol. 14:302–308.

Kepler, C. B. & Scott, J. M. 1981: Reducing bird count variability by training observers. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:366–371. Cooper Ornithological Society.

Ramsey, F. L. & Scott, J. M. 1981: Tests of hearing ability. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:341–345. Cooper Ornithological Society.

Raphael, M. G. 1987: Estimating relative abundance of forest birds: Simple versus adjusted counts. — Wilson Bull. 99:125–131.

Reynolds, R. T., Scott, J. M. & Nussbaum, R. A. 1980: A variable circular-plot method for estimating bird numbers. — Condor 82:309–313.

Salwasser, H., Hamilton, C. K., Krohn, W. B., Lipscomb, J. F. & Thomas, C. H. 1983: Monitoring wildlife and fish: mandates and their implications. — Trans. North Amer. Wildl. and Nat. Resources Conf. 48:297–307.

Scott, J. M., Mountainspring, S., Ramsey, F. L. & Kepler, C. B. 1986: Forest bird communities of the Hawaiian Islands: Their dynamics, ecology, and conservation. — Studies in Avian Biology 9. Cooper Ornithological Society.

Scott, J. M., Ramsey, F. L. & Kepler, C. B. 1981: Distance estimation as a variable in estimating bird numbers from vocalizations. — In: Ralph, C. J. & Scott, J. M. (eds.), Estimating numbers of terrestrial birds, Studies in Avian Biology 6:334–340. Cooper Ornithological Society.

Svensson, S. 1977: Land use planning and bird census work with particular reference to the application of the point sampling method. — Pol. Ecol. Stud. 3:99–117.

Verner, J. 1983: An integrated system for monitoring wildlife on the Sierra National Forest. — Trans. North Amer. Wildl. and Nat Resources Conf. 48:355–366.

—"— 1985: Assessment of counting techniques. — In: Johnston, R. F. (ed.), Current Ornithology, 2:247–302. Plenum Press, New York, NY.

Verner, J. & Ritter, L. V. 1985: A comparison of transects and point counts in oak-pine woodlands of California. — Condor 87:47–68.

—"— 1988: A comparison of transects and spot mapping in oak-pine woodlands of California. — Condor 90: 401–419.

Received 30.VIII.1987
Printed 22.XII.1989