OBSERVER DIFFERENCES IN THE NORTH AMERICAN BREEDING BIRD SURVEY

JOHN R. SAUER,[1] BRUCE G. PETERJOHN,[2] AND WILLIAM A. LINK[1]

[1]*Patuxent Wildlife Research Center and* [2]*Office of Migratory Bird Management,*
U.S. Fish and Wildlife Service, Laurel, Maryland 20708, USA

ABSTRACT.—Because count data collected in many bird surveys are only an index to population size, factors that can influence the counts must be identified and incorporated into analyses. Observer quality is often ignored in analyses of population changes from survey data, but observers differ in methods and capabilities and, hence, tend to count different numbers of birds. We assess the consequences of between-observer differences in counts for estimation of population trends in the North American Breeding Bird Survey. Observer differences in numbers of birds counted were found in 50% of the 369 species we examined. For many species, observers in later years tended to count more birds than observers in earlier years, suggesting an increase in observer quality over time. Analysis of population trends from 1966 through 1991 indicates that failure to include observers as covariables in the analysis results in an overly optimistic view of population trends. *Received 8 March 1993, accepted 24 October 1993.*

THE NORTH AMERICAN Breeding Bird Survey (BBS) is a primary source of systematically collected information on population changes for many species of North American birds. The BBS has been used to document changes in bird populations, such as those associated with severe winters (Sauer and Droege 1990a), population trends (consistent change in population size over time), and other population fluctuations (Robbins et al. 1986).

Estimation of population trends from BBS data can be controversial because count data collected by BBS observers tend to be imprecise, and any factor that influences these counts can potentially introduce spurious trends if not incorporated into the analysis (Sauer and Droege 1990b). Count data are subject to bias because the counts are indexes to population size, representing an unknown proportion of the actual number of birds present. The proportion can be small (<0.5) and is well known to exhibit intra- and interspecific variation (Barker and Sauer 1992). Although some sources of variability in the proportion are controlled by restricting the timing of the survey and weather conditions under which the survey can be conducted (Robbins et al. 1986), there are some factors that cannot be controlled.

One important factor that influences the proportion of birds counted is the ability of the observer to count birds. From the inception of the BBS and other roadside surveys, there has been evidence that observers differ greatly in their ability to detect birds (Robbins et al. 1986,

Sauer and Bortner 1990). Aside from obvious differences related to knowledge and hearing abilities of observers, many additional factors cause observers to differ in their counts on BBS routes. Each observer has a distinct strategy for counting birds, and all observers have difficulty counting the total number of birds when many individuals are singing at a stop (Bart and Schoultz 1984). In roadside surveys, the actual survey area may differ among observers because odometers vary and stops may not be shown accurately on maps; hence, different observers on the same survey routes often choose slightly different stop locations. Finally, some observers survey the routes by themselves, while for others an assistant records data. The availability of a recorder allows the observer significantly more time to concentrate on counting birds.

Although a single observer surveys a BBS route each year, most routes have had several observers, each of whom has surveyed the route for varying numbers of years. Of the 2,293 routes surveyed more than five years between 1966 and 1991, only 367 were surveyed by a single observer. For the other 1,926 routes, between 2 and 11 observers have collected data during at least one year (Fig. 1). Observation of individual BBS routes indicates that consistent changes in population level can be associated with changes in observers (Fig. 2).

Although observer differences may exist at the level of individual routes, they might only increase the variability of trend estimates without introducing bias into composite analyses in

which data from many routes are combined. However, if current BBS observers are more skilled than previous observers, observer changes will tend to increase counts on routes, thus creating a positive bias in estimates of population trends.

Because of the importance of unbiased estimation of trends for the management of bird populations, we quantified the extent of observer differences in the BBS. We first determined the extent of observer differences for counts of each species and for total number of species seen on BBS survey routes, and then we evaluated the bias in trend estimates that neglect to account for changes in interobserver quality over time.

METHODS

BBS data set.—We used BBS data from the interval 1966 through 1991 to document observer differences in counts of birds. The BBS, begun in 1966, is a roadside survey composed of more than 3,000 survey routes distributed across the United States and Canada (Robbins et al. 1986, Droege and Sauer 1990). Observers survey these routes one time each year, and count all birds heard and seen during 3-min listening periods at 50 stops spaced along the 39.4-km route. Attempts are made to have the same observers survey routes for series of years. BBS coordinators review the total number of bird species and individuals counted each year, and have informally adopted a code system for the acceptability of data. Route data are not used for years in which the route survey was incomplete, run under unacceptable conditions (such as bad weather), or when a new observer had large discrepancies in counts from those of previous observers. In this analysis, we use only data that passed this initial screening. We note, however, that the screening has not been consistent over time and among regions; hence, observers judged to provide acceptable data can differ in ability.

We evaluate two sets of dependent variables. The first, yearly total number of species observed on individual BBS routes, is used by BBS coordinators to assess observer quality and is a measure of regional species richness (Robbins et al. 1986). We evaluated these data for observer differences. The second data set consists of the yearly count data on individual BBS routes for each species, and has been the primary data base for estimation of population trends. We examined count data of each species for observer differences.

Analysis.—All of our analyses evaluated observer differences on individual routes. In our discussion of these analyses, the terms *observer differences* and *increases in observer quality* refer only to differences in counts among observers, not to changes in counting
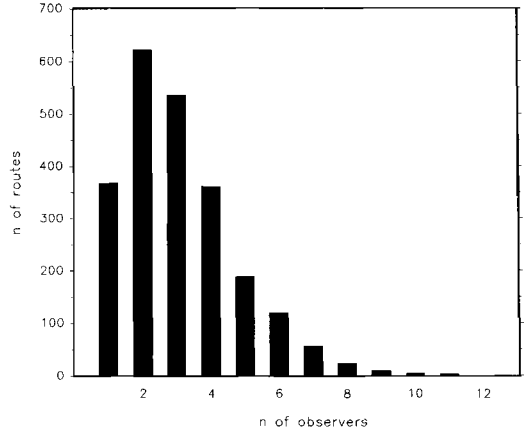

Fig. 1.   Number of BBS routes that have had specified number of observers. To appear in this histogram, route must have been surveyed for more than five years.

abilities over time for an individual observer. Changes in the capabilities of individuals will be addressed elsewhere (W. L. Kendall and B. G. Peterjohn unpubl. data).

Constraints on BBS analyses.—BBS data consist of time-series of counts from many survey routes. They are incomplete because many routes are not surveyed every year. Analysis of trends from these data is a challenging task, and some controversy exists regarding appropriate methods (e.g. Sauer and Droege 1990b). In an analysis of population trends, population changes are confounded with observer changes (i.e. changes in population counts coincident with changes in observers cannot be interpreted as trends in the population). Second, it is likely that population counts within BBS routes over time are autocorrelated. This autocorrelation can bias standard hypothesis tests for linear models (Draper and Smith 1981). Finally, routes differ greatly in both number of years surveyed and number of observer changes. Because of these difficulties, we chose to evaluate observer differences using several statistical methods, each of which has advantages under certain conditions.

Initial evaluation of observer differences.—We first examined the general null hypothesis that no differences existed among observers on a route. Because of the potential for autocorrelation, we did not use a linear model to examine the significance of observer covariables on individual routes, as was done in Sauer and Bortner (1990). Instead, we chose to examine differences in population counts between successive years on individual survey routes, grouped into two categories: (1) differences in counts when the route was surveyed by different observers; and (2) differences when the same observer surveyed the route for both years. To make the changes proportional, we log-transformed the data (adding 0.5 to avoid domain
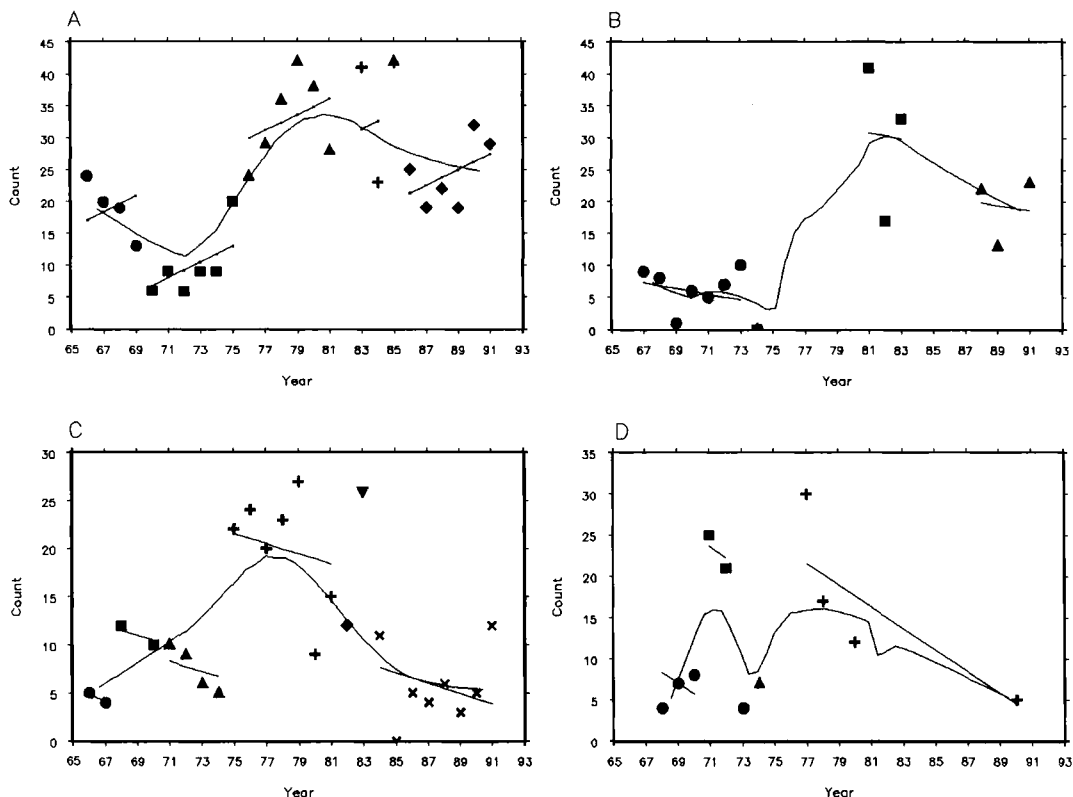
Fig. 2. BBS data for Red-eyed Vireo (*Vireo olivaceus*) from survey routes (A) 46008, (B) 65014, (C) 46011, and (D) 65002. For each route, we indicate each observer with a separate symbol, and include a common regression line with different intercepts for each observer (straight-line segments, used as route trend estimate in route-regression analysis) and a LOWESS-smoothed line (curved line; James et al. 1990).

errors) before the differences were calculated. The resulting data for a route surveyed for N years are the $N - 1$ differences.

Using these differences, we examined *Null Hypothesis A*: Population changes are of the same absolute magnitude regardless of whether observers change on the route.

Because we were interested only in whether changes in observers were associated with extreme population changes between years, we examined only the absolute magnitude of the change in counts. To test this hypothesis, we used three approaches. First, an extremely conservative approach used route data as replicates. On each route, we randomly selected one difference (count in year $t + 1$ − count in year t) from each category and determined if the difference in counts between years when the same observer ran the route was smaller than the difference in counts when different observers surveyed the route. If so, we counted the route as a *success*. We used a z-test to determine whether the percentage of successes on all routes differed from 50%. This binomial test has the advantage of mitigating consequences of within-route autocorrelation.

Alternatively, a Mann-Whitney test could be used for each route to compare the differences for the two groups, and the ranks could be summed among routes to provide a composite z-test for the species. This approach, although having greater power than the other tests, may be influenced by autocorrelation within routes, which could increase the probability of a type I error of the Mann-Whitney tests (Serfling 1968, Gastwirth and Rubin 1971). We conducted both this standard Mann-Whitney test and a modified version of the binomial test. This modified test involved randomly selecting one difference from each route and using a Mann-Whitney test (Lehmann 1975) on the randomly selected differences from all the routes on which a species occurred. The random selection Mann-Whitney test also produced z-tests of significance, and is unaffected by autocorrelation within routes. For all analyses, we used only routes that had at least one observer change and were surveyed for more than five years.

Evaluation of direction of observer changes.—Observer differences are particularly troublesome if they are directional, such that new observers are consistently better (or worse) than the observers they replace. To

evaluate whether new observers have increased or decreased counts of numbers of birds and species detected on BBS routes, we conducted two analyses that evaluated slightly different hypotheses about observer differences.

Linear model test for differences in observer effects. — We first took the approach of estimating observer effects of individual BBS routes using linear models, and determined whether these observer effects were more likely to be positive in later years. In this model, we use linear regression of natural logarithm of counts (plus the 0.5 constant) on time. Because observers change over time and differ in the number of birds they count, it is reasonable to incorporate observers as covariables, which allows the estimation of a common slope for an interval but different intercepts for each observer (Geissler and Sauer 1990). The equation for this regression on a route is:

$$\ln(C_{yr} + 0.5) = b_0 \cdot yr + \sum_{i=1}^{n} (b_i O_i), \qquad (1)$$

where C_{yr} is the count in year yr, b_0 is the slope (or trend) associated with time, O_i is an indicator variable for observer i in which $O_i = 1$ if observer i surveyed the route in a year and $O_i = 0$ if observer i did not survey the route, n is the number of observers, and b_i is the intercept associated with observer i. Each observer perceives the same trend, but the observer covariable shifts the trend line to accommodate differences in count levels for each observer.

With this model, we estimate observer effects on individual routes at specific times. Using routes as replicates, this approach can be modified to determine if later observers were better on BBS routes. The quality of an observer j on a route can be measured relative to other observers on the same route by

$$\Delta b_j = (b_j - \bar{b}), \qquad (2)$$

where $\bar{b}$ is the average of all the observer intercepts. Now, if observer quality is not associated with time, the size of Δb_j should not be associated with the first year observer j ran the route. Increases in Δb_j associated with observers that surveyed the route in later years provide evidence that recent observers tend to count more birds than earlier observers. This analysis assumes that trend is adequately modeled by a multiplicative slope parameter, and the additional differences in counts are associated with the observers. Consequently, gaps in coverage associated with missing years are adequately incorporated in this analysis.

Because of autocorrelation, direct comparisons of Δb_j for observers within a route may not be appropriate. We therefore chose to conduct this analysis using routes as replicates, and randomly selected one observer for each route (with associated data on the first year the observer surveyed the route) and estimated Δb_j for that observer. We measured the correlation between the first year the observer surveyed the route and Δb_j. This analysis addressed *Null Hy-*

pothesis B: No association exists between the relative efficiency of the observer (Δb_j) and the year in which the observer began to survey the route. We conducted this difference in intercept analysis for each route on which the species occurred and observers changed.

Directional tests using differences in counts. — The three procedures used in testing Null Hypothesis A can be modified to address whether observer differences tend to be directional, and evaluate *Null Hypothesis C*: Population changes are of the same magnitude and direction regardless of whether observers change on the route. We used the tests described above to examine this hypothesis, but left the signs on the differences. With the signed values, a significant positive z-statistic indicates that the differences are both significant and directional, and that observers are actually improving over time. We will refer to these results as the directional binomial, directional random-selection Mann-Whitney, and directional standard Mann-Whitney tests.

Effects of missing years. — BBS data are incomplete in that routes are not necessarily surveyed each year. Gaps in coverage are often associated with observer changes, but can also interrupt a series of years when a route is surveyed by the same observer. Sometimes a gap in coverage will correspond to a large change in the level of the time series of counts (Fig. 2). Analysis of the 1,926 routes with more than five years of data indicates that the average gap when the same observer surveyed the route in both years was 1.08 years, but the average gap when observers changed was 2.67 years.

We include data in our primary analysis regardless of the size of the gap. Gaps exist in both samples, and when observers change we cannot attribute the associated change in counts to either actual change in population size or observer effects. However, if trends exist in the population, changes in counts could be correlated with size of the gap in coverage. To assess the importance of these gaps to the analysis, we also applied the binomial test and directional binomial test to data for each species, omitting changes in counts associated with gaps of more than one year.

Summarizing observer differences. — Because we evaluated observer differences for many species, we had to summarize our results in a way that both documented observer differences and showed how the results differed among methods. We also were interested in variation in observer differences as a function of population density and species characteristics.

Summary analysis by species groups. — We summarized the results for groups of species by determining the percentage of species with observer differences both for all species and for grassland birds, Neotropical migrants, short-distance migrants, and permanent residents. The migration-status groups have been of recent scientific interest (e.g. Robbins et al. 1989), and it has been observed that grassland species have been declining as a group (Droege and Sauer 1990). Hence,
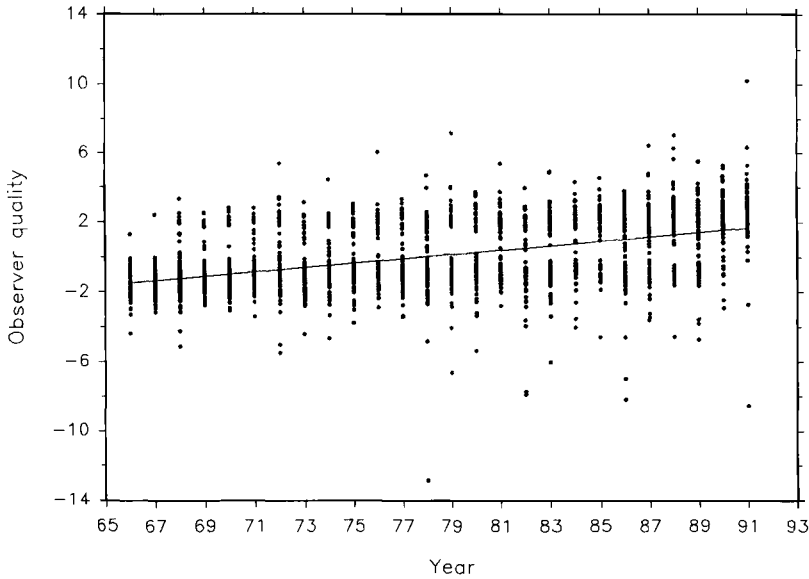
Fig. 3. Relative observer quality on BBS routes (as measured by difference between intercept term for observer and average observer on route) plotted against year of first survey by observer. Dependent variable for analysis was total number of species observed on BBS routes.

the composite effects of observer differences on our perception of trends in the groups are of interest. The species composition of each group is defined in Droege and Sauer (1990).

For each group, we can examine two percentages: (1) the percentage of species for which the null hypothesis is rejected (one-sided tests); and (2) the percentage of species that have signs that are consistent with rejection of the null hypothesis. Both percentages contain useful information, and the second percentage (which can be stated as the percentage of tests that reject at $\alpha = 0.50$) avoids the use of an arbitrary alpha level. These percentages can be tested to determine if they differ from those expected by chance alone. For example, we expect in 5% of the tests conducted to reject the null hypothesis by chance (at $\alpha = 0.05$); a value greater than 5% suggests more significant outcomes than expected by chance. Also, by chance, we expect to reject the null hypothesis 50% of the time when α is 0.50. Binomial tests were used to determine if observed percentages differ from those expected by chance under the null hypotheses. Note that species are used as replicates in these analyses and, therefore, we must assume that the species are independent trials for the estimation of the statistical significance.

Correlation with relative abundances and trends.—Evidence exists that variability in counts is associated with population density, and that species with low abundances tend to have high variability within routes, which might make observer differences difficult to detect. Population trends might also be confounded with observer effects. We estimated relative

abundance and population trends for each species, and correlated them with our measures of observer differences.

Evaluating consequences of observer differences in affecting trends.—It is possible to evaluate directly the consequences of not incorporating observer differences in BBS analyses. In methods using linear models such as in equation 1, the penalty for including observer covariables when no observer differences exist is decreased precision, because unneeded parameters are fit. However, not including observer covariables when they are important results in bias, because needed parameters are missing from the linear model. The route-regression method (Geissler and Sauer 1990, Sauer and Droege 1990b) can be used to estimate trends under the two conditions. The method consists of estimating a trend for each route based on equation 1, then estimating regional trends as weighted averages of these route trends. Variances of trends are estimated using bootstrapping of routes (Efron 1982). To estimate trends without observer effects we fit equation 1 without the observer covariables, then calculated the weighted average trends. We performed these analyses for the interval 1966-1991 for all species found on more than 20 BBS routes.

For each of the trend analyses, we estimated the percentage of species with increasing trends by group and for all species. We also calculated the difference in trends for the two analyses for each species, and determined the percentage of species for which the estimated trend without covariables was larger than the estimated trend with covariables by group and for all species. Consistent differences among the anal-

TABLE 1. Percentage of species with observer differences, as measured by changes between years on individual routes categorized as having either: (1) the same observers both years; or (2) different observers between years. First set documents observer differences without regard to directional changes in quality, and the second documents directional changes in observer quality. Information for binomial tests includes sample sizes, percentage of significant results, probability of hypothesis tests that the rate of significant results equalled 5%, percentage of results indicating greater differences in magnitude when observers changed, and probability value of hypothesis tests that the percentage of results greater than zero equalled 50%. Information also presented for binomial results when all gaps of more than one year are eliminated from analysis.

| | Binomial test | | | | | | | |
| | All data | | | | | No gaps | | |
Group	n	Percent signifi-cant	P	Percent > 0	P	n	Percent signifi-cant	Percent > 0
	Observer-difference tests (nondirectional)							
Grassland-nesting species	19	73.7	0.00	94.7	0.00	19	57.9	84.2
Permanent residents	53	47.2	0.00	88.7	0.00	52	40.4	78.8
Neotropical migrants	99	68.7	0.00	97.0	0.00	97	62.9	91.8
Short-distance migrants	99	60.6	0.00	93.9	0.00	94	53.2	79.8
All	369	49.6	0.00	84.6	0.00	341	44.3	78.9
	Observer-difference tests (directional)							
Grassland-nesting species	19	00.0	0.76	36.8	0.36	19	26.3	73.7
Permanent residents	53	17.0	0.00	50.9	1.00	52	17.3	63.5
Neotropical migrants	99	17.2	0.00	56.6	0.23	98	7.1	59.2
Short-distance migrants	99	20.2	0.00	55.6	0.32	96	15.6	64.6
All	369	14.4	0.00	57.7	0.00	346	11.6	61.8

yses would provide evidence that observer covariables are needed for unbiased estimation of trends.

RESULTS

Observer effects on observed total number of species.—Using the binomial test, we determined that the absolute magnitudes of randomly selected differences in total number of species recorded when observers changed were greater than differences when the observer remained the same on 63.6% of routes, which is significantly greater than 50% ($z = 12.32$, $n = 1,908$, $P < 0.00001$). The random-selection Mann-Whitney test showed a similar pattern, which resulted in rejection of the null hypothesis of similar differences with $P < 0.0001$ ($z = 9.16$). Highly significant results ($z = 24.83$, $P < 0.00001$) also were obtained from the standard Mann-Whitney test. For species totals, we therefore reject Null Hypothesis A.

The difference in intercept analysis indicated a significant increase in observer quality over time. The correlation of observer quality with year was 0.503 ($n = 1,911$, $P < 0.0001$), allowing us to reject Null Hypothesis B for total number of species observed. A plot of this relation shows a trend over time of improving observer quality (Fig. 3).

The binomial test indicated that the total number of species recorded on routes increased more when observers changed on 53.6% of routes, which is significantly greater than 50% ($z = 3.14$, $n = 1,911$, $P < 0.01$). The null hypothesis of similar differences also was rejected based both on the directional random-selection Mann-Whitney test ($z = 3.55$, $P < 0.01$) and the directional Mann-Whitney test ($z = 5.99$, $P < 0.001$), suggesting that Null Hypothesis C is false for species totals.

Observer effects on counts of individual species.— Binomial test results indicated significant observer effects in 49.6% of the 369 species used in the analysis (Table 1). Changes in counts were larger when observers differed between years for 84.6% of the species, which is much larger than 50% ($P < 0.0001$). Similar percentages of larger differences in counts when observers changed occurred in the random-selection Mann-Whitney tests (75.6%) and the standard Mann-Whitney tests (87.0%); among-species results of these tests were highly correlated with those of the binomial test ($r = 0.889$, $P < 0.0001$, and $r = 0.916$, $P < 0.0001$, respectively), although variation in actual significance levels of individual species occurs among results because of the random selection procedures used in the

TABLE 2. Percentage of species with observer differences, as measured by difference-in-intercept test using all data. Information presented for number of species analyzed, percentage of significant results, probability of hypothesis that rate of significant results equalled 5%, percentage of results indicating greater differences in magnitude when observers changed, and probability value for hypothesis test that percentage of results greater than zero equalled 50%. Information also presented for binomial results when all gaps of more than one year are eliminated from analysis.

Group	n	Difference-in-intercept test			
		Percent significant	P	Percent > 0	P
Grassland-nesting species	19	68.4	0.00	84.2	0.00
Permanent residents	53	32.1	0.00	62.3	0.10
Neotropical migrants	99	28.3	0.00	61.6	0.03
Short-distance migrants	99	33.3	0.00	62.6	0.02
All	369	26.3	0.00	59.6	0.00

tests. Consequently, Null Hypothesis A is rejected for at least 50% of the species we analyzed.

Elimination of gaps in coverage of more than one year tends to decrease the sample sizes for each species by approximately 30%, and decreases the number of species in the analysis to 341 (Table 1). However, the percentage of species with significant observer effects (44.3%) and the percentage of species in which changes were greater in magnitude when observers changed (78.9%) were only slightly smaller than those described above.

Correlations of relative observer efficiency with the first year the observer surveyed the route indicate that for 26.3% of the 369 species there were significant increases in observer quality over time (Table 2). Overall, for 59.6% of the species there were positive correlations of observer quality and time. Both of these percentages are much larger than those expected by chance ($P < 0.0001$), and Null Hypothesis B is therefore rejected for a significant number of species. We list the species with significant correlations in the Appendix.

Directional binomial test results indicate that 14.4% of species had significant positive observer effects, and 57.7% had positive observer effects (Table 1). Both of these results are greater than expected due to chance ($P < 0.004$). As above, similar percentages resulted from the random-selection Mann-Whitney test (62.3%) and standard Mann-Whitney test (65.0%); their species-specific results were highly correlated with those of the binomial test ($r = 0.714, P < 0.0001$, and $r = 0.652, P < 0.0001$, respectively). These results indicate that Null Hypothesis C is rejected for significantly more bird species than would be expected by chance.

The elimination of gaps of more than one year again decreased sample sizes and slightly lowered the percentage of species with significant results (11.6%). The percentage of species with test statistics greater than zero increased slightly, to 61.8%. Both of these percentages are greater than those expected by chance ($P < 0.01$).

Analysis by groups.—Because the random-selection Mann-Whitney and standard Mann-Whitney tests results were highly correlated with the binomial test results for individual species, we discuss only the binomial test results for group analyses.

Importance of observer effects differed greatly among groups of birds. Grassland and Neotropical-migrant species included more than 65% of species with significant observer effects (Table 1). Omitting gaps of more than one year from the analysis resulted in similar, but generally slightly smaller, percentages of species with observer differences.

For both the difference in intercept (Table 2) and the directional binomial tests (Table 1), all three migration-status groups had notably large percentages of directional observer effects. The difference in intercept analysis showed generally higher percentages of significant results for each group than did the binomial test. In grassland species, for example, no species were significant in the binomial test, but 68% were significant in the difference in intercept test (Table 2).

Omitting gaps of more than one year from the directional binomial test of observer differences again resulted in generally lower percentages of significant results. However, for grassland species, elimination of the gaps resulted in a much higher percentage of significant directional observer differences (Table 1).

TABLE 3. Percentage of species with increasing populations (estimated trend greater than zero), by group and for all species, for analyses not including observers as a covariable (not included) and including observers as a covariable (included). Data also presented on percentage of species in each group for which trend estimated without observers was larger than trend estimated with observers (percent greater), along with number of species in the group for which trends were estimated.

| | Observer covariable | | Percent greater | n (species) |
Group	Included	Not included		
Grassland nesting species	23.8	28.6	76.2	21
Permanent residents	49.2	61.0	74.6	59
Neotropical migrants	58.3	73.3	74.8	103
Short-distance migrants	46.6	64.7	77.7	103
All	52.1	66.7	72.6	409

Correlation of relative-abundance and observer differences.—We calculated the correlation of relative abundance of species with the observer-difference test results for 377 species. Relative abundances of species were positively correlated with z-statistics for the binomial ($r = 0.538$, $P < 0.0001$), the random-selection Mann-Whitney ($r = 0.501$, $P < 0.0001$), and the standard Mann-Whitney ($r = 0.622$, $P < 0.0001$) nondirectional-observer-difference tests. For the directional-observer-difference tests, the correlations were not significant for the binomial ($r = -0.015, P = 0.772$), the random-selection Mann-Whitney ($r = 0.045$, $P < 0.388$), or the standard Mann-Whitney ($r = 0.029$, $P < 0.576$) tests, but there was a positive association between the difference in intercept analysis results and relative abundance ($r = 0.268$, $P < 0.001$).

Correlation of trend and observer differences.—Population trend of bird species was negatively associated with the z-test values for nondirectional-observer-difference tests (binomial, $r = -0.180$, $P < 0.0005$; random-selection Mann-Whitney, $r = -0.126, P = 0.0159$; standard Mann-Whitney, $r = -0.166$, $P = 0.0014$) and with correlations from the difference in intercept analysis for directional observer differences ($r = -0.270, P < 0.0001$), but not with z-test values from the other directional-observer-change tests.

We partitioned the data into two sets corresponding to species with estimated trends greater than zero, and estimated trends less than zero to determine if the correlations were consistent. The correlation of trend and magnitude of observer differences was negative for increasing species (binomial test, $r = -0.265, P < 0.001$) in each test, but positive for declining species (binomial test, $r = 0.156$, $P < 0.001$), suggesting observer differences are most evident in populations that are not experiencing extreme trends. Splitting the difference in intercept correlation into increasing and declining groups caused the correlations to be not significant (increasing trends, $r = -0.099, P > 0.18$, decreasing trends, $r = -0.084, P > 0.20$).

Agreement among measures of directional observer differences.—The results among species for the difference in intercept test were not highly correlated with either the directional binomial test ($r = 0.012, P > 0.8$) or the other directional tests (all $P > 0.1$). This reflects the variability induced by the random-selection procedures and the differences between the methods.

Comparison of trend results.—Trends fit without observer covariables tend to indicate increasing populations much more frequently than trends estimated with observer covariables (Table 3). Overall, when observer covariables are explicitly incorporated into trend analysis, 52.1% of the species had positive point estimates of trend. Omitting the observer covariable results in 66.7% of species with positive estimates of trend. This result is consistent across groups, although the magnitude differs greatly among the groups. Analysis of differences among species shows that trends without covariables are more positive for 72.6% of species (Table 3).

DISCUSSION

We have demonstrated that changes in observers on BBS routes can greatly influence our perception of population changes on the routes. We have documented observer effects on counts in the majority of species surveyed by the BBS, and directional observer differences occur in a significant number of species. The observer differences are most apparent in species seen in large numbers of BBS routes, and extreme trends
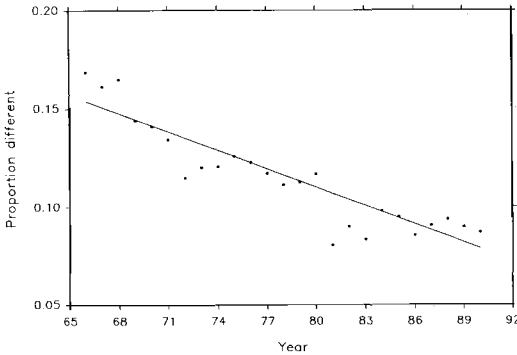
Fig. 4. Proportion of BBS routes run by different observers in two years, given that route was run in both years. Proportions indexed by first year of two-year period.

tend to make observer differences less apparent. Failure to include observer differences as cov-ariables in trend analyses causes a positive bias in estimated trends. These results are the first quantitative documentation of observer effects on BBS count data.

Evaluating observer quality from existing survey data.—We judge that it is critically important to view the problem of observer differences from several different perspectives because of poten-tial statistical difficulties involved with any par-ticular approach. As with the analysis of pop-ulation trends from the data, the assumptions that are made for each method can have a great effect on the estimators. For example, the dif-ference in intercept analysis checks for differ-ences in the level of the population associated with observers. Because this approach estimates observer quality using all data for each observ-er, it may be more efficient in that it avoids the problems associated with aberrant counts in the first (or last) year an observer surveys a route (e.g. the learning effect documented by Erskine 1978). Therefore, although tests of Null Hy-potheses B and C both examine whether ob-server changes tend to produce higher counts, tests of Null Hypothesis B evaluate differences in the mean level of each observer in the context of trend, but tests of Null Hypothesis C directly evaluate changes in counts between years.

Of course, in any statistical analysis it is im-portant to consider deficiencies that could un-dermine the credibility of the results. In our analysis, we must determine if it is possible that there are factors other than observer differences that could cause us to reject the null hypothesis of no directional observer effects. Both the tem-

poral pattern of observer changes and the trend in the population could play a role in affecting our view of the importance of observer differ-ences. For example, if counts on a route are postulated to fit log-linear models with year and observer as random variables that have vari-ances that are functions of the means, the dif-ferences in counts between years may not be symmetrically distributed when a dispropor-tionate number of observer changes occur late in the time series. If this occurred, it could lead to a slightly higher probability that differences associated with observer changes are larger than those differences not associated with observer changes. We examined the proportion of routes on which observers changed from year to year, and found that the proportion decreased over time (Fig. 4).

Although the temporal pattern of observer changes is probably not affecting our results, we feel that the redundancy in the statistical tests provides confidence in our results. How-ever, the random procedures we used to cir-cumvent within-route autocorrelation intro-duce variability in results for individual species. Therefore, we have also analyzed patterns at the level of groups in this paper. Consistency of the composite results for groups increases our confidence in the applicability of the results.

The difference in intercept analysis evaluates observer differences in terms of mean levels of each observer in the context of within-route trends, thereby avoiding difficulties associated with both aberrant counts by new observers and variability in the data. This analysis contains a higher percentage of directional observer dif-ferences, indicating a greater power to detect differences. Also, the results correlate negative-ly with trends, but do not correlate with results of the other directional observer methods, in-dicating that removal of trends from the anal-ysis may provide more sensitivity for detecting observer differences. We believe that the dif-ference in intercept analysis provides the clear-est indications of systematic differences among observers over time in the BBS.

Why should observers be improving over time?— The BBS was initiated in 1966, after a pilot year in which experimental surveys were conducted in Maryland and Delaware (Robbins et al. 1986). By 1968, survey routes were established throughout the continental United States and southern Canada, and new routes have been initiated each subsequent year. The survey cre-

ated an unprecedented need for observers who could consistently identify birds by song, and during the early years of the survey there were difficulties in finding qualified observers for all routes. Most new observers did not have experience with procedures for counting birds, and all observers in the first years were learning the techniques of the BBS. Therefore, any low-count effect (Erskine 1978) resulting from observer inexperience would have occurred on all routes during only the first years of the survey. Later observer changes did not occur simultaneously on large numbers of routes.

We also speculate that the observer-quality screening conducted by the BBS coordinators tends to improve the quality of observers. After an observer has surveyed a route, data exist that can be used to judge the qualifications of subsequent observers. New observers that count fewer birds or miss species are likely to be disqualified, and the data from the earlier observer provides the context for assessing the skill of later observers. It is likely that new observers of poorer quality will be viewed with caution, but observers of equal or better quality will be accepted without question. Therefore, over time the new observers may tend to be at least equal in competence to earlier observers.

Observer differences and relative abundances of species.—Species that are counted in large numbers on BBS routes tend to have greater observer differences. Empirical studies have shown that observers have difficulty in distinguishing numbers of birds singing when even relatively small numbers are present (Bart and Schoultz 1984). When birds occur in flocks, time constraints often preclude accurate counts, and observers count groups differently; hence, consistency should occur only within individual observers. Relatively small changes in stop locations, which may occur when observers change on routes, often can obscure views of habitats (such as barns) where flocks of birds exist, and cause additional observer differences in counts.

Observer differences and trends.—Observer differences tend to be less apparent in species that have undergone extreme trends in population size. A change in counts between years on a BBS route has two components, a signal representing the actual change in population, and noise representing the variability associated with the percentage of birds not counted. If the signal forms an important part of the count

changes, it is unlikely that the Mann-Whitney and binomial tests presented in this paper would be sensitive to changes associated with observer differences. The correlation results demonstrate the pattern of extreme trends associated with an inability to detect observer differences.

Ramifications for analysis of BBS data.—It is well known that regional and temporal changes in survey intensity can affect the results of simple analyses of BBS data. Geissler and Noon (1981), for example, documented that simple averages of route data over time can bias trends when new routes are initiated that contain more birds than the older routes. Our results suggest that even more caution is necessary in BBS analyses to avoid biases in estimation of population parameters. Observer differences complicate every aspect of the analysis of BBS data.

Observer effects and trend analysis.—Our perceptions of extreme changes in bird populations probably are little affected by observer differences. Any reasonable method of analysis can detect whether birds have undergone a precipitous change in population status. However, subtle changes in bird populations often are of great interest to ecologists, and we have demonstrated that, in the BBS, observer differences can greatly change our perceptions of trends in bird populations.

Observer differences complicate analysis of BBS data at several levels of regional aggregation. On the level of individual routes, the time series is not just incomplete because of missing years, but the mean level of the series can change with observers. Estimation of trends at the level of individual routes, therefore, is likely to be biased, as will any within-route smoothing procedure (Sauer and Droege 1990b, James et al. 1990).

The bias also exists at the regional scale. Because trends estimated without observer covariables tend to be more positive than trends estimated with observer covariables (Table 3), it cannot be argued that observer differences are only a random factor that disappears in regional analyses that combine route data. Analysis of trend, autocorrelation, and other time-series components are all affected by observer differences, and searches for patterns in BBS data are best conducted within data collected by individual observers.

Observer effects and biodiversity analyses.—We have documented observer effects in total number of species of birds seen on BBS routes. This

result is a direct consequence of the observer effects at the level of individual species because it reflects the increased probability of detecting any species present along the route. Naive analysis of BBS data could indicate an increase in bird biodiversity associated with increases in observer quality. All surveys that collect index data are subject to the possibility of confounding changes in observer quality or effort with changes in population or community parameters. We suggest that any analysis of species richness from count data use statistical procedures available for estimation of richness (Bunge and Fitzpatrick 1993).

Observer effects and mapping of bird distributions from BBS data.—One important application of BBS data has been in defining existing breeding ranges and patterns of relative abundance of bird species. Use of computer-mapping procedures is popular for making maps from BBS data (e.g. Sauer and Droege 1990a). These results suggest that use of maps to evaluate short-term changes in bird distributions could provide misleading results as a consequence of improvement in observer quality. In particular, comparisons of maps from different time intervals must be made with caution because increases in observer quality will be confounded with population increase in these maps.

Observer effects and data checking.—One enduring technical question among BBS coordinators has been quality control of BBS data. The results presented here suggest that all criteria for evaluating observer quality using existing route data are arbitrary, as most observers will differ in at least some component of their survey of a route. If observer information is incorporated into the analysis, it becomes less important to insure a high level of consistency among observers, and the real issue becomes whether each observer is consistent in his or her counts over time. Unfortunately, there is evidence that individual observers change in quality over time, and this topic clearly merits further investigation.

LITERATURE CITED

BARKER, R. J., AND J. R. SAUER. 1992. Modelling population change from time series data. Pages 182-194 in Wildlife 2001: Populations (D. R. McCullough and R. H. Barrett, Eds.). Elsevier Applied Science, New York.

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

BUNGE, J., AND M. FITZPATRICK. 1993. Estimating the number of species: A review. J. Am. Stat. Soc. 88: 364-373.

DRAPER, N. R., AND H. SMITH. 1981. Applied regression analysis. John Wiley and Sons, New York.

DROEGE, S., AND J. R. SAUER. 1990. North American Breeding Bird Survey annual summary 1989. U.S. Fish Wildl. Serv., Biol. Rep. 90:1-22.

EFRON, B. 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial Applications of Mathematics, Philadelphia, Pennsylvania.

ERSKINE, A. J. 1978. The first ten years of the cooperative breeding bird survey in Canada. Can. Wildl. Serv. Rep. Ser. 42:1-61.

GASTWIRTH, J. L., AND H. RUBIN. 1971. Effect of dependence on level of some one-sample tests. J. Am. Stat. Assoc. 66:816-820.

GEISSLER, P. H., AND B. R. NOON. 1981. Estimates of avian population trends from the North American Breeding Bird Survey. Pages 42-51 in Estimating the numbers of terrestrial birds (C. J. Ralph and J. M. Scott, Eds.). Stud. Avian. Biol. 6.

GEISSLER, P. H., AND J. R. SAUER. 1990. Topics in route regression analysis. Pages 54-57 in Survey designs and statistical methods for the estimation of avian population trends (J. R. Sauer and S. Droege, Eds.). U.S. Fish. Wildl. Serv., Biol. Rep. 90(1).

JAMES, F. C., C. E. McCULLOCH, AND L. E. WOLFE. 1990. Methodological issues in the estimation of trends in bird populations with an example: The Pine Warbler. Pages 84-97 in Survey designs and statistical methods for the estimation of avian population trends (J. R. Sauer and S. Droege, Eds.). U.S. Fish. Wildl. Serv., Biol. Rep. 90(1).

LEHMANN, E. L. 1975. Nonparametrics: Statistical methods based on ranks. Holden Day, San Francisco.

ROBBINS, C. S., D. BYSTRAK, AND P. H. GEISSLER. 1986. The Breeding Bird Survey: Its first fifteen years, 1965-1979. U.S. Fish. Wildl. Serv., Resour. Publ. 157.

ROBBINS, C. S., J. R. SAUER, R. S. GREENBERG, AND S. DROEGE. 1989. Population declines in North American birds that migrate to the Neotropics. Proc. Natl. Acad. Sci, USA 86:7658-7662.

SAUER, J. R., AND J. B. BORTNER. 1990. Population trends from the American Woodcock singing-

ground survey, 1970–88. J. Wildl. Manage. 55: 300–312.

SAUER, J. R., AND S. DROEGE. 1990a. Recent population trends of the Eastern Bluebird. Wilson Bull. 102:239–252.

SAUER, J. R., AND S. DROEGE. 1990b. Survey designs and statistical methods for the estimation of avian population trends. U.S. Fish Wildl. Serv., Biol. Rep. 90(1).

SERFLING, R. J. 1968. The Wilcoxon two-sample statistic on strongly-mixing processes. Ann. Math. Stat. 39:1202–1209.

APPENDIX. Species with significant changes in observer quality over time, as indicated by significant correlation from difference in intercept analysis. To be included in list, a species must have had at least 21 BBS routes with acceptable data for estimating the correlation. We include the number of routes, the correlation, and the statistical significance of the null hypothesis that the correlation is equal to zero. Probabilities are one-sided; hence [a], $P < 0.05$ and [b], $P < 0.01$ indicate a significant positive observer effect. We also include species with very high probabilities ([c], $P > 0.95$, [d], $P > 0.99$), indicating a decline in observer quality.

Western Grebe, *Aechmophorus occidentalis*, 100, 0.229[a]; American Bittern, *Botaurus lentiginosus*, 550, 0.118[b]; Snowy Egret, *Egretta thula*, 205, −0.136[c]; Cattle Egret, *Bubulcus ibis*, 368, −0.106[c]; Green Heron, *Butorides virescens*, 1253, −0.065[c]; Glossy Ibis, *Plegadis falcinellus*, 44, 0.256[a]; Mute Swan, *Cygnus olor*, 27, 0.423[a]; Wood Duck, *Aix sponsa*, 1019, −0.059[c]; Mallard, *Anas platyrhynchos*, 1381, −0.118[d]; Northern Shoveler, *A. clypeata*, 242, −0.109[c]; Gadwall, *A. strepera*, 244, −0.114[c]; Redhead, *Aythya americana*, 161, −0.168[c]; Lesser Scaup, *A. affinis*, 176, 0.193[b]; Common Goldeneye, *Bucephala clangula*, 104, 0.241[b]; Ruddy Duck, *Oxyura jamaicensis*, 172, −0.173[c]; Black Vulture, *Coragyps atratus*, 413, −0.094[c]; White-tailed Kite, *Elanus leucurus*, 84, 0.232[a]; Broad-winged Hawk, *Buteo platypterus*, 730, −0.135[d]; Red-tailed Hawk, *B. jamaicensis*, 1813, −0.040[c]; Gray Partridge, *Perdix perdix*, 220, −0.163[d], Ring-necked Pheasant, *Phasianus colchicus*, 957, 0.081[b]; Northern Bobwhite, *Colinus virginianus*, 1017, 0.296[b]; Lesser Yellowlegs, *Tringa flavipes*, 62, 0.248[a]; Upland Sandpiper, *Bartramia longicauda*, 516, −0.084[c]; Long-billed Curlew, *Numenius americanus*, 149, 0.208[b]; Common Tern, *Sterna hirundo*, 137, 0.156[a]; Arctic Tern, *S. paradisaea*, 21, −0.578[d]; Band-tailed Pigeon, *Columba fasciata*, 155, 0.142[a]; Yellow-billed Cuckoo, *Coccyzus americanus*, 1252, 0.076[c]; Common Barn-Owl, *Tyto alba*, 68, 0.275[a]; Eastern Screech-Owl, *Otus asio*, 297, −0.101[c]; Western Screech-Owl, *O. kennicottii*, 35, 0.310[a]; Great Horned Owl, *Bubo virginianus*, 1197, −0.061[c]; Burrowing Owl, *Athene cunicularia*, 262, 0.112[a]; Long-eared Owl, *Asio otus*, 27, −0.439[c]; Lesser Nighthawk, *Chordeiles acutipennis*, 85, −0.285[d]; Whip-poor-will, *Caprimulgus vociferus*, 515, 0.130[b]; Black Swift, *Cypseloides niger*, 63, 0.223[a]; Chimney Swift, *Chaetura pelagica*, 1355, 0.069[b]; Ruby-throated Hummingbird, *Archilochus colubris*, 1111, −0.060[c]; Costa's Hummingbird, *Calypte costae*, 52, 0.296[a]; Belted Kingfisher, *Ceryle alcyon*, 1574, −0.048[c]; Red-headed Woodpecker, *Melanerpes erythrocephalus*, 964, 0.063[a]; Acorn Woodpecker, *M. formicivorus*, 123, 0.236[b]; Red-bellied Woodpecker, *M. carolinus*, 953, 0.067[a]; White-headed Woodpecker, *Picoides albolarvatus*, 52, −0.301[c]; Red-shafted and Yellow-shafted flicker, *Colaptes auratus*, 505, 0.074[a] and 1503, 0.192[b], respectively; Olive-sided Flycatcher, *Contopus borealis*, 552, 0.146[b]; Eastern Wood-Pewee, *C. virens*, 1291, 0.105[b]; Acadian Flycatcher, *Empidonax virescens*, 642, 0.077[a]; Least Flycatcher, *E. minimus*, 814, 0.064[a]; Dusky Flycatcher, *E. oberholseri*, 199, −0.202[d]; Eastern Phoebe, *Sayornis phoebe*, 1248, 0.104[b]; Eastern Kingbird, *Tyrannus tyrannus*, 1642, 0.064[b]; Scissor-tailed Flycatcher, *T. forficatus*, 209, 0.117[a]; Horned Lark, *Eremophila alpestris*, 1292, 0.123[b]; Bank Swallow, *Riparia riparia*, 1009, 0.054[a]; Barn Swallow, *Hirundo rustica*, 1957, 0.065[b]; Steller's Jay, *Cyanocitta stelleri*, 248, 0.163[b]; Blue Jay, *C. cristata*, 1462, 0.091[b]; Scrub Jay, *Aphelocoma californica*, 224, 0.193[b]; Black-billed Magpie, *Pica pica*, 379, 0.138[b]; American Crow, *Corvus brachyrhynchos*, 1860, 0.130[b]; Carolina Chickadee, *Parus carolinensis*, 648, 0.166[b]; Mountain Chickadee, *P. gambeli*, 223, 0.185[b]; Plain Titmouse, *P. inornatus*, 149, 0.159[a]; Tufted Titmouse, *P. bicolor*, 976, 0.096[b]; Red-breasted Nuthatch, *Sitta canadensis*, 635, −0.090[c]; Brown-headed Nuthatch, *S. pusilla*, 214, 0.263[b]; Cactus Wren, *Campylorhynchus brunneicapillus*, 99, 0.232[a]; Canyon Wren, *Catherpes mexicanus*, 139, 0.261[b]; Carolina Wren, *Thryothorus ludovicianus*, 813, 0.148[b]; Marsh Wren, *Cistothorus palustris*, 330, 0.122[a]; Blue-gray Gnatcatcher, *Polioptila caerulea*, 923, 0.096[b]; Black-tailed Gnatcatcher, *P. melanura*, 42, 0.311[a]; Eastern Bluebird, *Sialia sialis*, 1220, 0.097[b]; Townsend's Solitaire, *Myadestes townsendi*, 187, 0.136[a]; American Robin, *Turdus migratorius*, 1870, 0.124[b]; Varied Thrush, *Ixoreus naevia*, 125, 0.217[b]; Gray Catbird, *Dumetella carolinensis*, 1442, 0.115[b]; Northern Mockingbird, *Mimus polyglottos*, 1270, 0.066[b]; Brown Thrasher, *Toxostoma rufum*, 1414, 0.090[b]; Crissal Thrasher, *T. crissale*, 24, −0.494[d]; Sprague's Pipit, *Anthus spragueii*, 82, 0.224[a]; Phainopepla, *Phainopepla nitens*, 82, 0.221[a]; Loggerhead Shrike, *Lanius ludovicianus*, 1039, 0.131[b], European Starling, *Sturnus vulgaris*, 1985, 0.095[b]; White-eyed Vireo, *Vireo griseus*, 701, 0.076[a]; Yellow-throated Vireo, *V. flavifrons*, 917, 0.080[b]; Red-eyed Vireo, *V. olivaceus*, 1464, 0.049[a]; Golden-winged Warbler, *Vermivora chrysoptera*, 234, 0.125[a]; Orange-crowned Warbler, *V. celata*, 283, −0.119[c]; Northern Parula, *Parula americana*, 707, −0.064[c]; Yellow Warbler, *Dendroica petechia*, 1584, 0.047[a]; Chestnut-sided Warbler, *D. pensylvanica*, 546, 0.084[a]; Prairie Warbler, *D. discolor*, 585, 0.141[b]; Palm Warbler, *D. palmarum*, 64, −0.306[d]; Ovenbird, *Seiurus aurocapillus*, 909, 0.085[b]; Common Yellowthroat, *Geothlypis trichas*, 1716, 0.104[b], Hooded Warbler, *Wilsonia citrina*, 446, 0.107[a]; Yellow-breasted Chat, *Icteria virens*, 951, 0.067[a]; Summer Tanager, *Piranga rubra*, 585, 0.072[a]; Northern Cardinal, *Cardinalis cardinalis*, 1203, 0.105[b]; Rose-breasted Grosbeak, *Pheucticus ludovicianus*, 843, −0.058[b]; Indigo Bunting, *Passerina*

APPENDIX.   Continued.

cyanea, 1299, 0.127[b]; Painted Bunting, *P. ciris*, 217, 0.233[b]; Dickcissel, *Spiza americana*, 650, 0.187[b]; Rufous-sided Towhee, *Pipilo erythrophthalmus*, 1452, 0.097[b]; Chipping Sparrow, *Spizella passerina*, 1695, 0.086[b]; Field Sparrow, *S. pusilla*, 1220, 0.219[b]; Vesper Sparrow, *Pooecetes gramineus*, 1075, 0.075[b]; Lark Bunting, *Calamospiza melanocorys*, 230, 0.136[a]; Savannah Sparrow, *Passerculus sandwichensis*, 1054, 0.078[b]; Grasshopper Sparrow, *Ammodramus savannarum*, 1126, 0.081[b]; LeConte's Sparrow, *A. leconteii*, 100, 0.242[b]; Song Sparrow, *Melospiza melodia*, 1523, 0.133[b]; White-throated Sparrow, *Zonotrichia albicollis*, 431, 0.180[b], Slate-colored and Oregon junco, *Junco hyemalis*, 385, 0.148[b] and 269, 0.140[a], respectively; Chestnut-collared. Longspur, *Calcarius ornatus*, 99, 0.230[a]; Bobolink, *Dolichonyx oryzivorus*, 829, 0.094[b]; Red-winged Blackbird, *Agelaius phoeniceus*, 1991, 0.198[b]; Eastern Meadowlark, *Sturnella magna*, 1320, 0.248[b]; Western Meadowlark, *S. neglecta*, 977, 0.199[b]; Common Grackle, *Quiscalus quiscula*, 1573, 0.167[b]; Brown-headed Cowbird, *Molothrus ater*, 2003, 0.164[b]; American Goldfinch, *Carduelis tristis*, 1575, 0.122[b]; House Sparrow, *Passer domesticus*, 1878, 0.139[b].