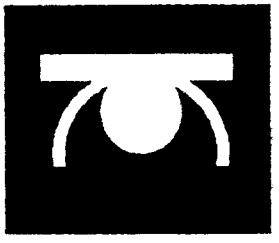
ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

ILLINOIS NATURAL HISTORY SURVEY



EVALUATION OF THE UPPER MISSISSIPPI RIVER CORRIDOR MONITORING PROGRAM FOR THE MIGRATORY BIRDS

Submitted to the U.S. Fish and Wildlife Service by:

Jeffrey D. Brawn
Sarah Hartung
Illinois Natural History Survey
Center for Wildlife Ecology
Champaign, IL

September 18, 1996

TABLE OF CONTENTS

INTRODUCTION

The U. S. Fish and Wildlife Service (USFWS) has undertaken an avian monitoring program as part of a comprehensive effort to conserve nongame birds in areas such as those belonging to the National Wildlife Refuge System. Recent interest in the conservation of neotropical migrant landbirds (NMLBs) has led to further need and coordinated monitoring efforts between the USFWS and other groups participating in the Partners in Flight programs. The primary goals of the USFWS program are to monitor population trends, determine the status of species over time, from local to regional spatial scales, and to assess avian habitat associations.

Beginning in FY 1994, the USFWS Region 3 adopted point counts as the method for monitoring NMLBs (USFWS 1994). Since that time, much sampling has been carried out and, for certain districts and field stations, habitat measurements have also been taken. An especially critical area for NMLBs within Region 3 is the Upper Mississippi River corridor (UMR) which includes the Mark Twain NWR (MTNWR) and the Upper Mississippi River NW&FR (here referred to simply as UMRNW&FR). In 1995, USFWS personnel contracted the Illinois Natural History Survey to evaluate the effectiveness of the monitoring program in the Upper Mississippi River Corridor.

The evaluation of the monitoring program focused on the following objectives and topics:

I) To determine the effectiveness of the program in meeting its objectives and make recommendations for the enhancement of a standardized protocol that will to assess avian biodiversity with the UMR and to assess bird-habitat associations within the corridor.

Specific considerations include: 1) recommendations for specific field method(s) to estimate avian species richness and relative abundances of birds during the breeding season and during migrations periods; 2) Recommendations for a specific sampling design using the methods identified in (1). This element will consider the effects of scale and the hierarchical nature of the management "units" within the UMR. Specific issues include the number of sampling points to use at different scales or strata , the number of visits / point / season, and a protocol for selecting points.

II) To recommend procedures for analyses of monitoring data. This objective includes the following elements: 1) Recommendations of methods to detect trends in species richness or relative abundances. Trends would be either over time or among sampling areas. Variation over space will include comparisons within the UMR (say, between habitats) and between the UMR and areas or regions outside the UMR. 2) Recommendations for assessing relationships between variation in species richness/abundance and variation in habitat structure (breeding and migration periods). 3) Recommendations for data management.

III) To suggest research or additional monitoring needs that might enhance understanding of avian biodiversity in the UMR.

EVALUATION OF THE UMR MONITORING PROGRAM

Overview of point counts and evaluation of monitoring methods used in the UMR

Considerable effort has been devoted to establishing standardized monitoring methods for nongame birds (e.g., Ralph et al. 1995). Many methods have been proven

effective and, often, the specific objectives of the program will dictate which field technique is most appropriate (see Bibby et al. 1992, Butcher et al. 1993). At present, the point count method (Hutto et al. 1986) is becoming the standard field technique for assessing species composition and estimating relative abundances of NMLBs.

The major advantages of point counts (fixed radius, unlimited-radius, or otherwise) are that all types of birds (understory, canopy, etc.) are counted, the method is portable to all types of habitats or seasons, and it is an efficient for encountering rare species (Butcher et al. 1993). Another advantage is that point counts are a cost effective monitoring method (Butcher et al. 1992). The major disadvantage of point counts during the breeding season is that the viability of populations on the areas or habitats being censused cannot be determined. Specifically, the reproductive status of birds detected on a point is unknown. Whether males heard on a count are on territories with mates or unmated "floaters" cannot be determined in most cases. This is an especially large drawback when attempting to establish habitat associations because the habitat around singing perches may be different than that for nesting and foraging . Moreover, point counts do not yield information on important demographic parameters such a survival rates, recruitment, and nest success (including rates of predation and brood parasitism). This issue is discussed further in Recommendations. Another consideration of point counts is that the potential for observer error or variability is high (Ralph et al. 1995). This variability highlights a potentially serious problem common to point counts and other techniques for estimating abundances where counts are incomplete and based on indices (e.g., number of registrations / point). If detection probabilities (i.e., probability that an individual will be detected and recorded given presence in area) vary among species, habitats, observers, years, or whatever, then

serious statistical biases can be introduced that can affect the validity of inferential statistical procedures (Lancia et al. 1994, Barker and Sauer 1995, Pendleton 1995). This problem will also be considered in Recommendations.

The specific protocol for point counts varies according to duration of the count, the radius within which birds are counted, and the number of times each point is visited within a given season. The basic protocol currently used within the UMR also differs by refuge and habitat. At the UMRNW&FR in open habitats, counts are 10 minute counts of birds detected out to a fixed radius of 100 m. Subtotals by distance (0 to 50 m and 50 to 100 m) and time (up to 5 min and 5 to 10 min) are also recorded. In more forested habitats, birds are recorded out to 50m with 0-25m and 25-50m subtotals. At the MTNWR, birds are recorded at fixed distances up to 50m and those beyond this distance as well (note that the databases supplied for preparation of this report did not always carry subtotals as specified above). The recommended number of visits to each point is 1 / season for the breeding season and up to 4 / season during migration. The number of points visited has varied widely among refuges, districts, etc. and among years. Depending on refuge or district, habitat measurements have been made at the census points using the basic protocol recommended by Ralph et al. (1993).

Evaluation of the UMR monitoring program involved analyses of extant data for certain questions and assessment given current recommendations/state of knowledge for programs using point counts. A major consideration was the adequacy of sample sizes for characterizing species composition, estimating relative abundances of particular species, and for detecting differences in these quantities among years, habitats, or other management units such as refuges.

To assess the adequacy of sample sizes to date and make recommendations for changes (if any), several analyses were carried out. Rarefaction analyses were carried out to assess sample sizes and species composition. Rarefaction is a technique that assess the relationship between the number of individuals observed and the expected number of species in that sample. Often rarefaction is used to compare species richness among two samples where sampling effort differs (Gotelli and Graves 1996). Here, rarefaction was used (via the algorithm supplied by Ludwig and Reynolds 1988) simply to assess the point or asymptote (if any) beyond which increased sampling would likely not lead to the expectation of more species. Rarefaction analyses presented here consider only the data from UMRNW&FR because results from this refuge and the MTNWR were similar and led to similar conclusions.

A second technique employed was to use resampling methods to obtain bootstrap estimates of species abundances. Resampling is a relatively new technique for analysis when the validity of "classical" hypothesis testing or sampling properties are uncertain. For this, randomly drawn subsets of samples (points) of different sizes (n = 10, 25, 50, points etc.) were drawn with replacement from the sample for a given year (e.g., MTNWR Wapello District for 1995). The mean number of registrations/ point in these subsets (via 1000 draws) was then calculated and precision of the estimates assessed using bootstrap generated confidence intervals.

Finally, power analyses were used to assess effects of sample size and ability to detect differences as associated with questions about differences among habitats or variation over time. The two quantities estimated for this analyses were β and power. β is

defined as the probability of making a Type II error; that is, concluding that there are no real differences among, for example, population means when they actually do exist. Power is simply 1- β and is a measure of how "powerful" the sample is in detecting differences. Generally, power increases as sample size increases. The power analyses here were run on estimates of species richness and relative abundance for selected species (with emphasis on NMLBs). In most cases, variance estimates and "effect sizes" (see below) for power analyses were derived from collected data. In certain instances "generic" analyses using standard normal distributions and effect sizes were run.

To assess the protocol for each count (distance, timing, etc.),we compared different subtotals of the numbers of individuals and species observed. For example the number of species observed during the first five minutes was compared with that observed during the second five minutes. Evaluation and recommendations for selection of points also relied on the point-count literature and standard sampling protocols.

Analyses of bird-vegetation relationships were somewhat limited by the availability of data. Nonetheless, several techniques have been recommended.

For nearly all analyses, two sources of UMR data were used. First, data supplied by E. Nelson that cover the UMRNW&FR (primarily from the Winona District) for 1994 and 1995. Second, data supplied by J. Quinliven from the MTNWR (Wapello District) for 1993-1996. Sources of these data will be referred to simply as UMRNW&FR and MTNWR, respectively. Other data from the UMR were kindly supplied, but the above sources were deemed to be representative. Analyses of habitat effects were carried out exclusively on the data from UMRNW&FR, whereas annual variation was assessed with

data from MTNWR. Nearly all of the analyses outlined above were carried out for the breeding season as well as the two migration periods.

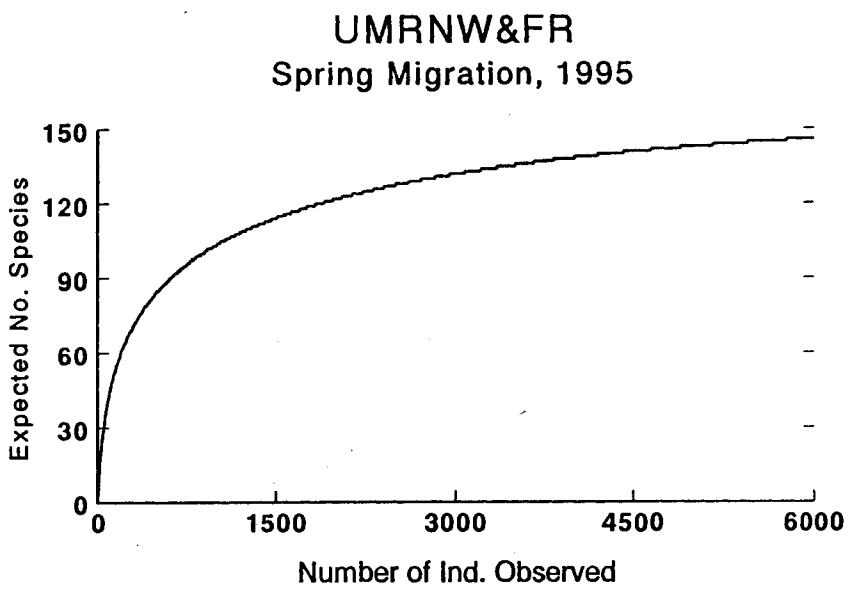
Sampling effort and the adequacy of sample sizes used in the UMR monitoring program

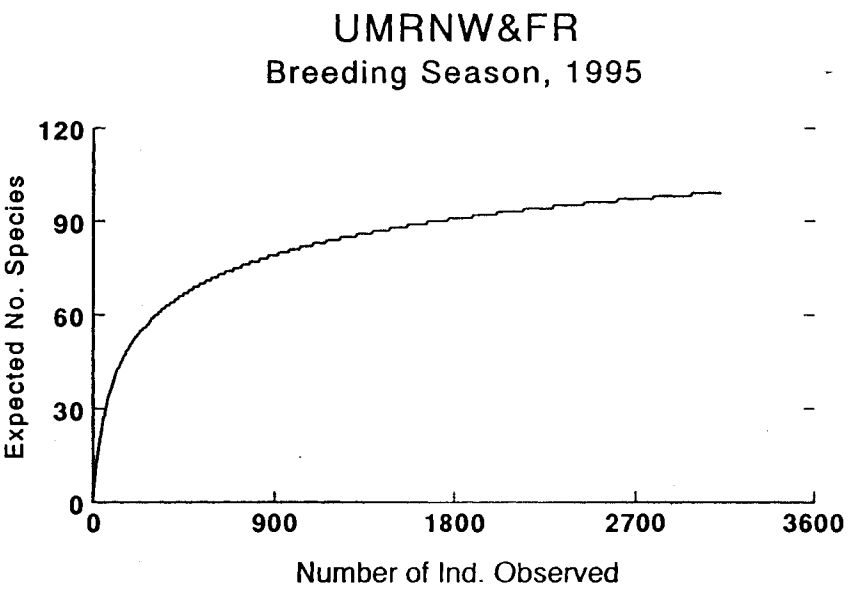
Sampling effort (see Table 1) varied widely among years and geographic locations (note that sample size here refers to the number of point counts carried out, not the number of different points - these quantities varied depending on the number of repeat visits to a given point within a season). For example, the number of point counts visited within the MTNWR during the fall migration varied from 31 in 1995 to 307 in 1993 (Table 1). Varying sampling effort further supported the use of rarefaction to assess the sufficiency of sampling effort. Most data available from the UMRNW&FR were collected in 1995.

Rarefaction analyses - Rarefaction analyses were carried out for the three sampling periods for UMRNW&FR data pooled over all habitats and separately for each habitat designation. For the fall migration and breeding season samples over all habitats (again, data from E. Nelson for the Winona District), it appears that the expected number of species does not increase appreciably once about 1500 to 2000 individuals have been observed (Fig. 1). Therefore, sampling effort beyond this level might not be efficient in terms of estimating species richness or composition. For both seasons, the asymptote was about 90 species. In contrast, increases in the expected number of species for the spring migration period did not level off (at about 140 species) until about 4000 individuals had been observed.

Table 1. Sample sizes (number of points) for UMR monitoring program databases.

Season	Year	MTNWR	UMRNW & FR
Spring	1993	122	-----
	1994	70	4
	1995	51	179
Breeding	1993	218	-----
	1994	40	86
	1995	20	150
Fall	1993	307	-----
	1994	60	-----
	1995	31	123





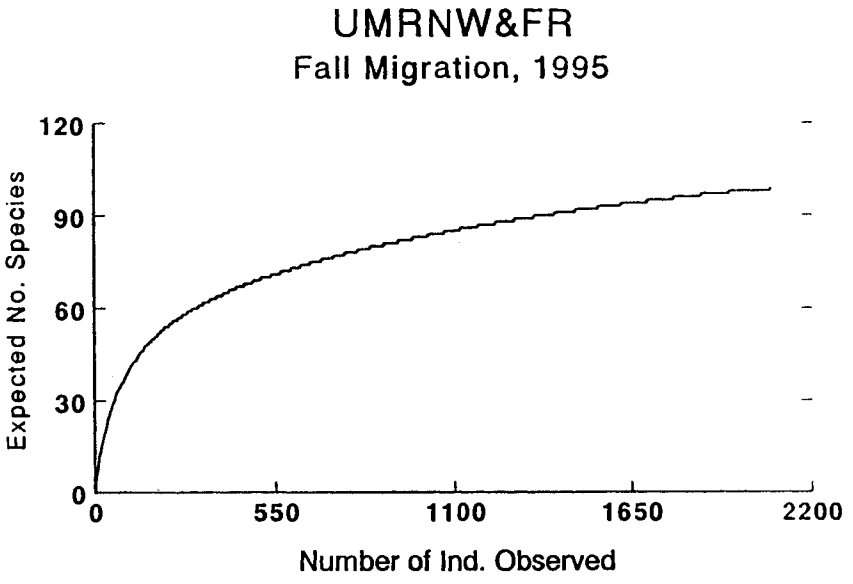


Fig. 1. Results of rarefaction analyses for species richness on UMRNW & FR during three sampling periods (all habitats pooled).

Note that in some cases the same points were visited more than one time within a season. Analyses indicated that these results were insensitive to single visit versus multiple visit samples.

Observed species richness during the spring migration period in the UMRNW&FR was considerably higher than that observed during the other sampling periods. Note that the expected number of species based on sampling effort and rarefaction analyses may be different than the number of species actually observed (see Ludwig and Reynolds, 1988). Interestingly, the total number of species actually observed in 1995 was considerably higher in the UMRNW&FR than in the 1993 MTNWR (Table 2, available data in the MTNWR were so few in 1995, that a meaningful within-year comparison with rarefaction was not possible). In contrast, the number of species observed / point was consistently greater on the MTNWR. Reasons for this pattern may be reflect true biology (e.g., species diversity may vary among points more in the UMRNW&FR) or be a sampling artifact owing to, for example, differential skills or numbers of observers.

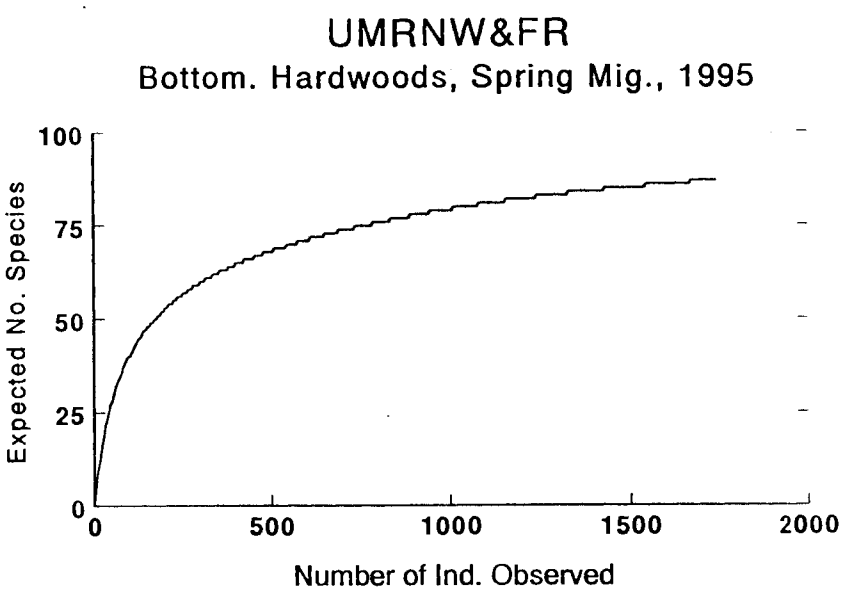
Observe species richness and needed sampling effort varied among habitats. For example, more species were observed in bottomland hardwood habitat than in upland prairie. In the breeding season, about 600 and 300 individuals would need to be observed in each habitat, respectively, to adequately characterize species composition (Fig. 2). Based on observed species richness (Table 3) and inspection of rarefaction curves, required sampling effort by habitat during the breeding season would rank as follows: bottomland hardwood > upland forest > upland prairie >emergent wetland > mixed wetland /upland. As was the case with all habitats pooled, for a given habitat type, more sampling is generally needed in the spring than in the breeding season or fall sampling periods.

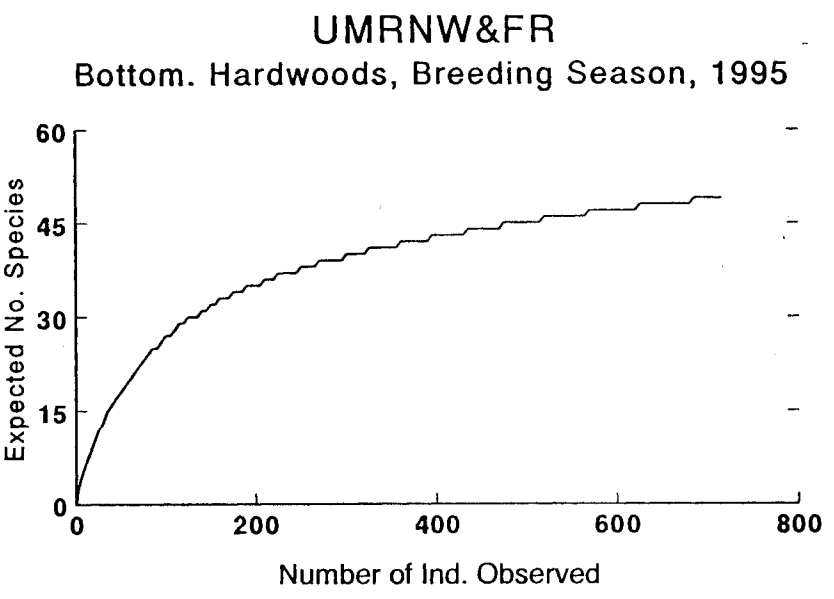
Table 2. Overview of Upper Mississippi River corridor bird sampling in two refuges.

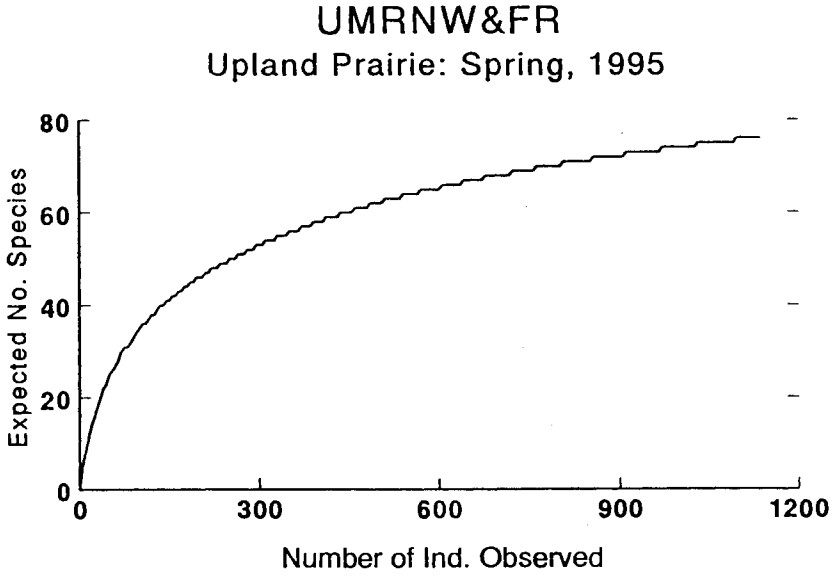
Location	Season	# of Species Observed	# of Species observed/point (x, [SE])
UMRNW & FR (1995)	Spring	148	12.1 (0.4)
	Breeding	109	12.9 (0.4)
	Fall	102	7.8 (.38)
MTNWR (1993)	Spring	70	21 (0.65)
	Breeding	86	17.9 (0.99)
	Fall	98	14.7 (0.96)

Table 3. Observed species richness and numbers of individuals by habitat on UMRNW & FR in 1995.

Habitat	Season	# Species Observed	# Individuals Observed
Bottomland Hardwood	Spring	88	1882
	Breeding	56	1314
	Fall	56	580
Emergent Wetland	Spring	60	1248
	Breeding	34	184
	Fall	44	420
Mixed Wetland/Upland	Spring	86	973
	Breeding	48	505
	Fall	50	314
Upland Forest	Spring	61	476
	Breeding	44	213
	Fall	33	218
Upland Prairie	Spring	78	1282
	Breeding	51	350
	Fall	45	463







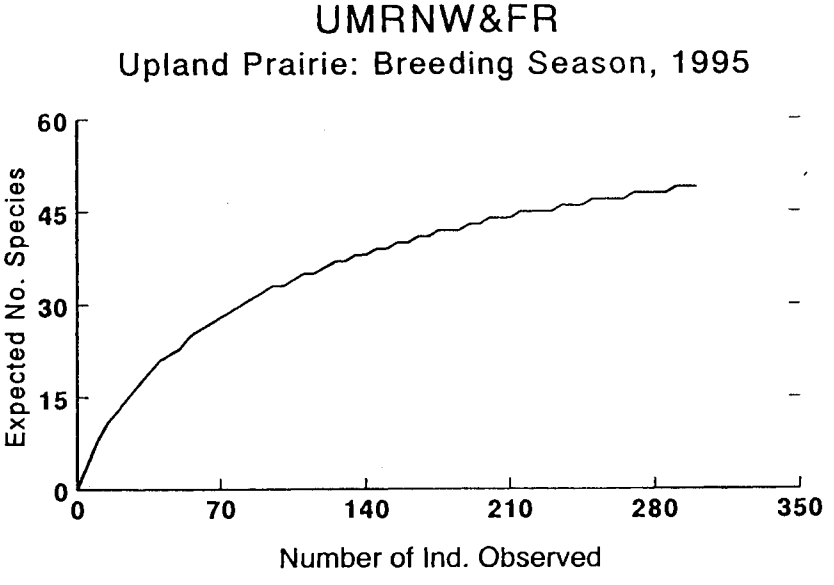
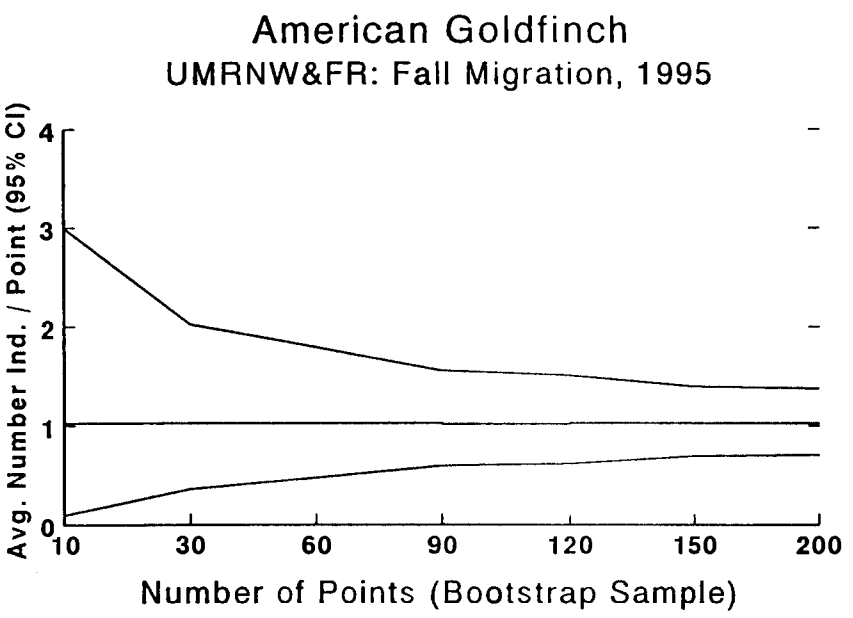


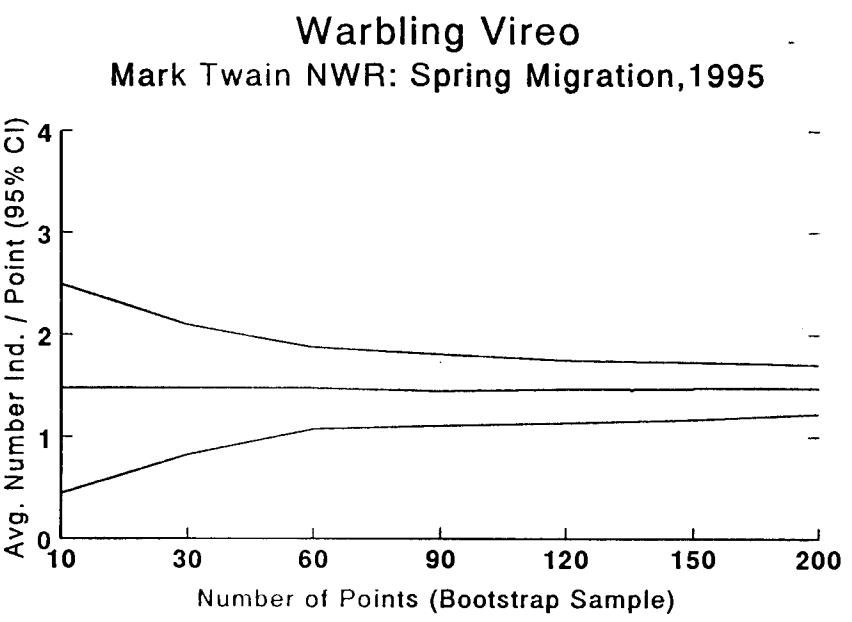
Fig. 2. Representative results of rarefaction analyses with specific habitats on the UMRNW & FR.

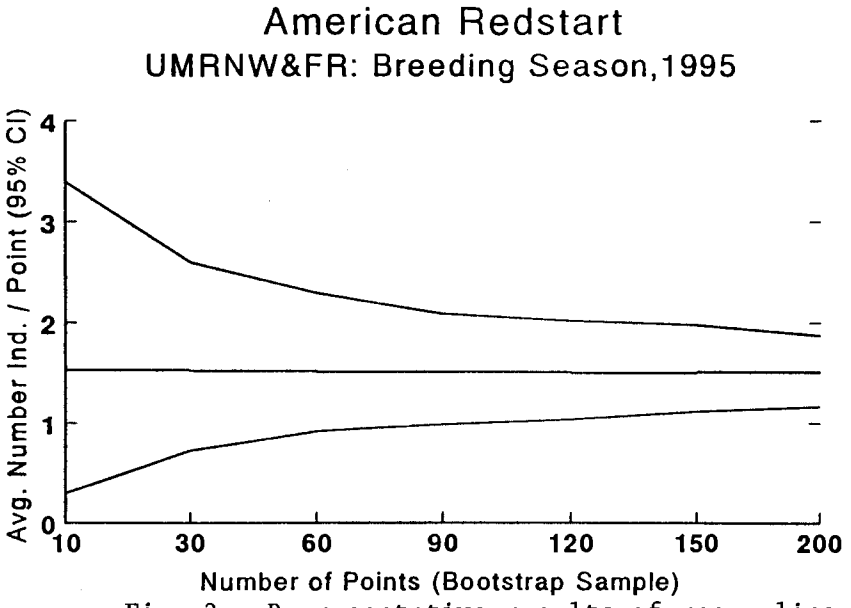
Recommendations for the number of points needed / habitat and for the MTNWR and UMRNW&FR, based on the average number of individuals observed / point and other analyses, are presented below.

Bootstrapping analyses of point counts - Bootstrapping or resampling of point count data were performed to assess where - in terms if sampling intensity - the width of confidence intervals stabilizes around estimated parameters such as abundances. The rationale here was that sampling should be sufficient to maximize precision of the estimates. Generally, if precision is higher than the power of statistical test is enhanced (note , however, that precision of an estimate does not guarantee its accuracy; i.e., the difference between the estimate of an parameter and the true population value of that parameter). To this end, bootstrap samples of different sizes (10 to 200) were drawn from populations of points (e.g., all points from the UMRNW&FR from a given year's fall census) and 95% confidence interval, based on variation among the samples, were constructed. For all runs, 1000 "draws" or subsamples were taken. Details about resampling and bootstrapping are found in Efron and Tibshirani (1993).

Bootstrapping was carried out for estimated abundances of selected species. Results of these analyses are illustrated by example in Fig. 3. For all habitat types pooled, the subsamples consistently converged in the observed parameter estimate. In other words, even for small (sub) sample sizes, average abundances were close to those derived from the observed full sample. Variation among samples was relatively high with samples sizes of 30 or less. Generally for sample sizes of > 75-100, the width of confidence intervals did not change appreciably. Therefore in terms of precision, additional sampling might not be needed for estimating relative abundances. This pattern held for all sampling periods







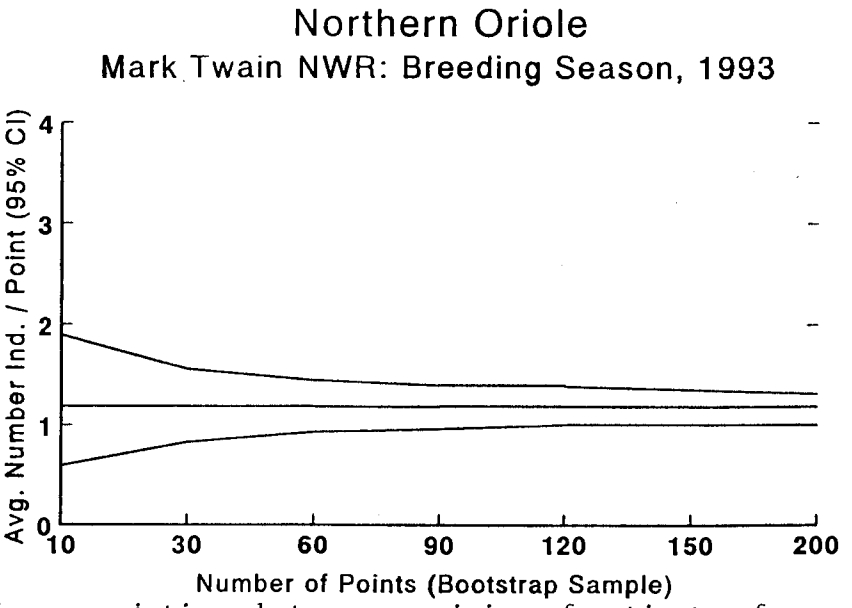


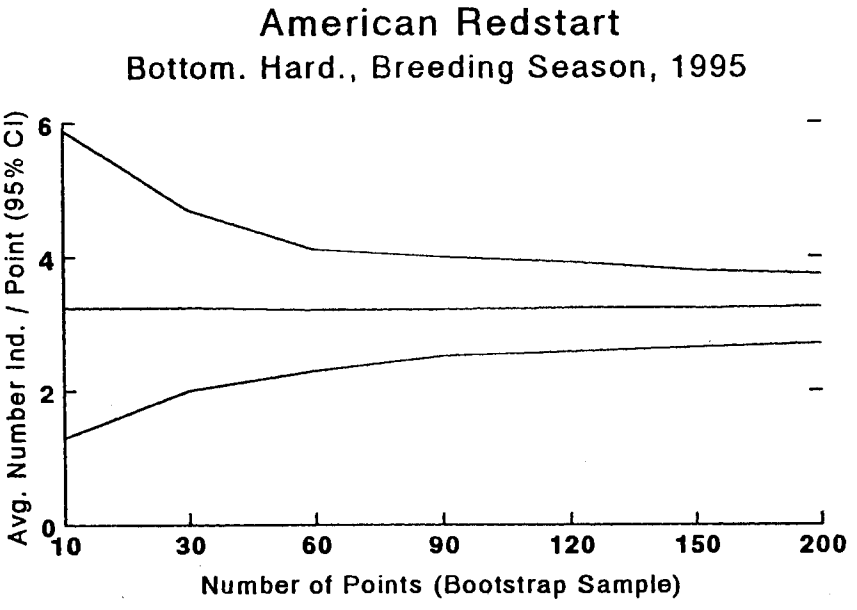
Fig. 3. Representative results of resampling analyses for associations between precision of estimates for abundance and sample size within specific habitats (UMRNW & FR). Outer lines depict (bootstrap) confidence intervals, center lines depict average estimated abundance.

although variance among samples tended to be greater for the spring and fall migration periods than the breeding season. In addition, variance among samples tended to be greater for relatively uncommon species.

For specific habitats, results were similar to those for all habitats combined - a sample size of 80 points or more yielded relatively precise estimates (Fig. 4). In general, precision was comparatively low for abundances of species in habitats where they were rare or uncommon.

Power Analyses - Several types of power analyses were run. For selected comparisons, we ran t-tests or ANOVAS on differences among refuges, habitats, or years and then calculated β and power. In most cases, we compared overall species richness and abundances of selected species. We also calculated β and power $(1 - \beta)$ over a range of sample sizes using variance estimates from observed data. Power is an important quantity because a monitoring program should be able to detect trends when, in fact, they are occurring. Otherwise, important changes from a management-conservation perspective might go unnoticed.

To assess power over a range of potential sample sizes, we follow Cohen (1988) and express power in light of different "effect" sizes. In everyday terms, "effect" is the magnitude of differences that are being compared. In ANOVA, for example, this would be the magnitude of differences among sample means from different habitats. Effects are expressed by the quantity "d" and not in terms of the original units. D can be loosely interpreted as a % difference in sample means or some other quantity. The quantity d can therefore be compared from study to study. We also follow Cohen and estimated power



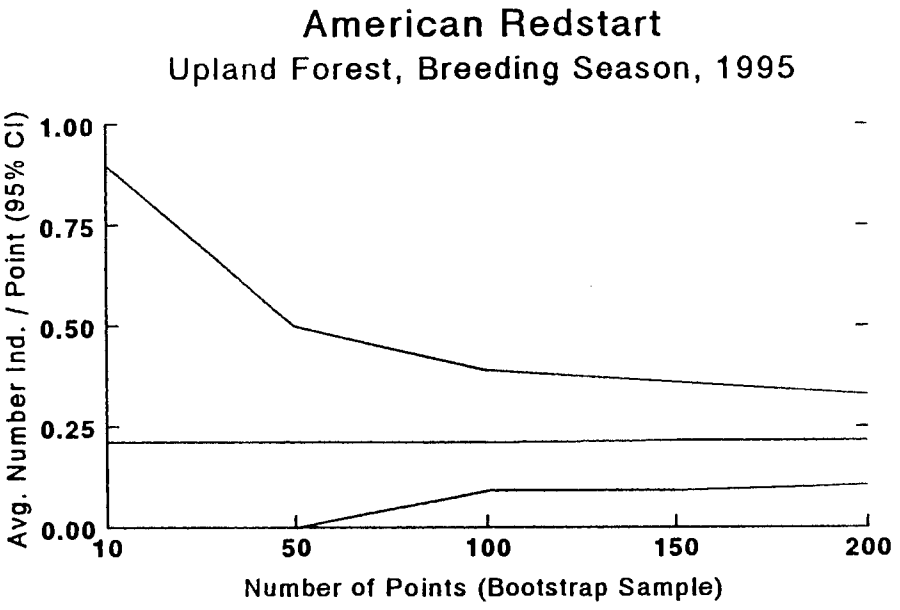


Fig. 4. Representative results of resampling analyses for associations between precision of estimates for abundance and sample size within specific habitats (UMRNW & FR). Outer lines depict 95% (bootstrap) confidence intervals, center lines depict average estimated abundance.

for small (d = .2), medium (d = .5), and large (d = .8) effects. To assess variation between or among years, we used data from the MTNWR (1993 to 1995). For habitat effects, we used the five most commonly visited habitats (bottomland hardwood, emergent wetland, mixed-wetland upland, upland forest, and upland prairie) and the 1995 data from the UMRNW&FR.

For power and comparisons at a large geographical scale, we assessed differences in specie richness and overall abundances (i.e., number of individuals observed / point) between the MTNWR and the UMRNW&FR. In 1995, estimated species richness was greater on the MTNWR than the UMRNW&FR within all three census periods (Fig. 5). These differences were highly significant (t-tests, $P < 0.01$) for all seasons. In term of effect size, these mean differences of species richness observed / point were over 0.7. With the sample effort expended (Fig. 5), especially on the UMRNW&FR, power to detect these differences was accordingly high (> .95 for all tests). Variation in estimated species richness between the MTNWR and the bottomland hardwood points in the UMRNW&FR were significant for all seasons (t-tests, $P < 0.01$).

Analyses of power to detect differences in species richness during the breeding season over a range of sample sizes for small medium and large effects is (Fig. 6) indicated that for large and medium effects, samples sizes of 150 or more points / refuge resulted in power of $\geq .80$. For small effects, samples sizes of 400 / refuge resulted in β of about .50. Power analyses for species richness during the spring and fall migration period were nearly identical.

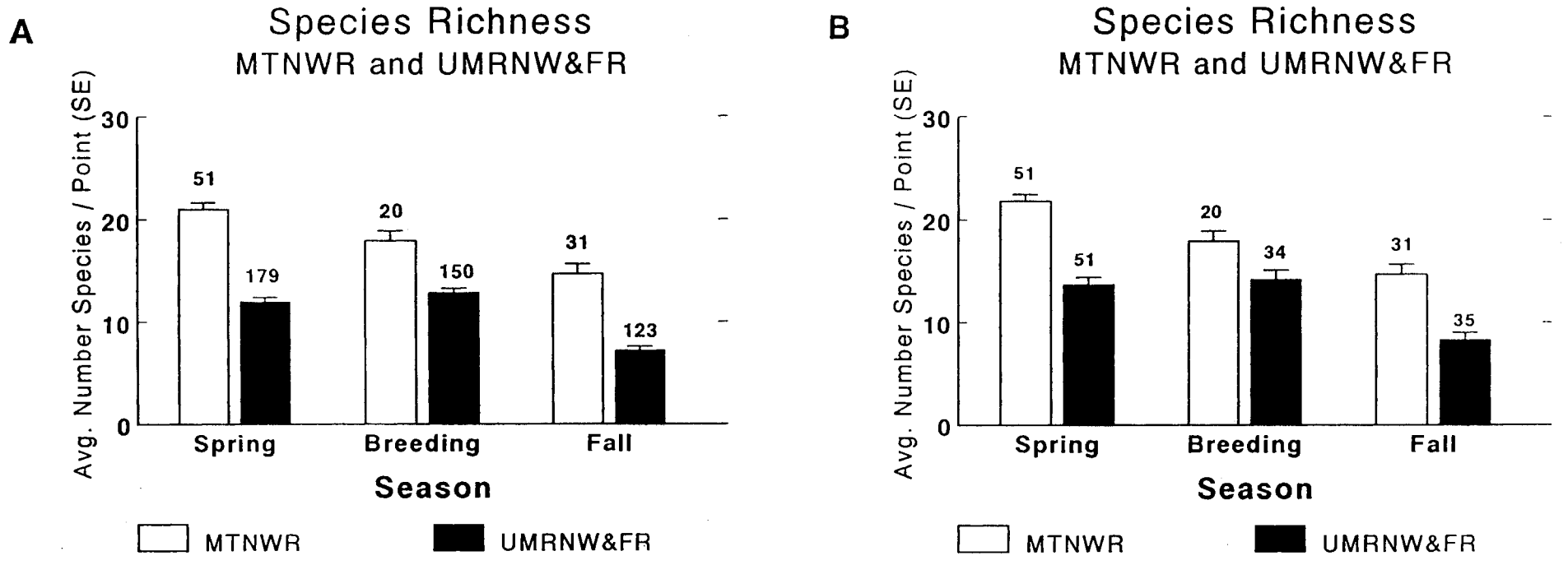


Fig. 5. Average number of species observed during three sampling periods on two refuges: a) all habitats; b) bottomland hardwood habitat only. Sample sizes are shown on tops of bars.

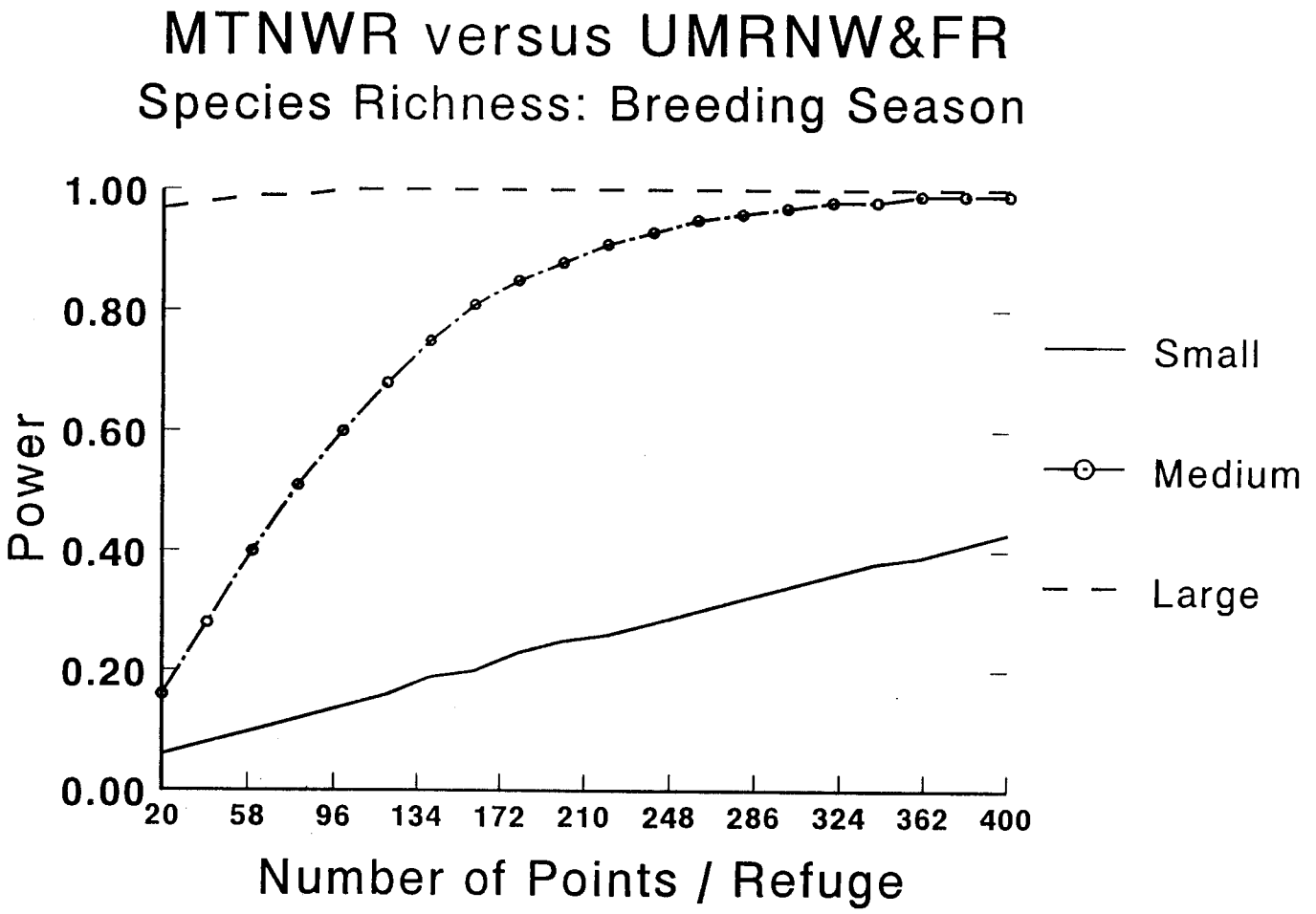


Fig. 6. Power analysis for detecting differences in species richness between refuges. Lines represent small, medium and large "effect" sizes and roughly correspond to 20%, 50%, and 80% differences, respectively.

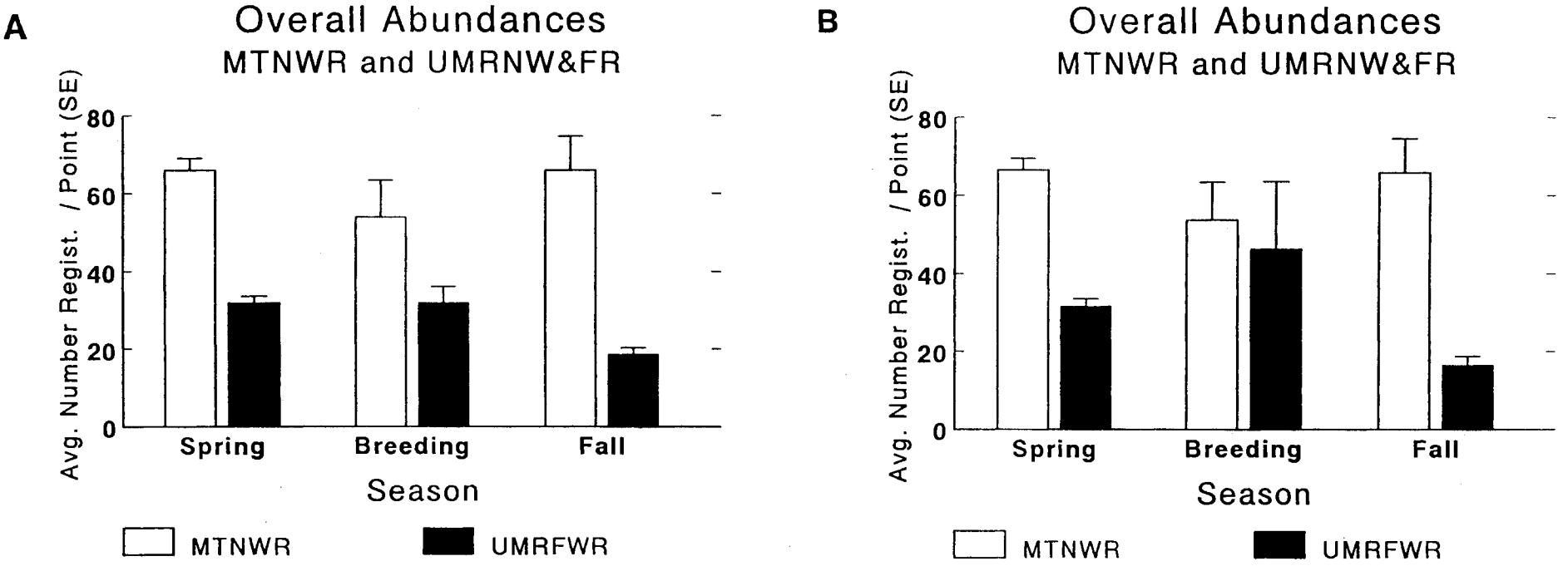


Fig. 7. Average number of individuals observed/point during three sampling periods on two refuges: a) all habitats; b) bottomland hardwood habitat only. Sample sizes are the same as those shown in Fig. 5.

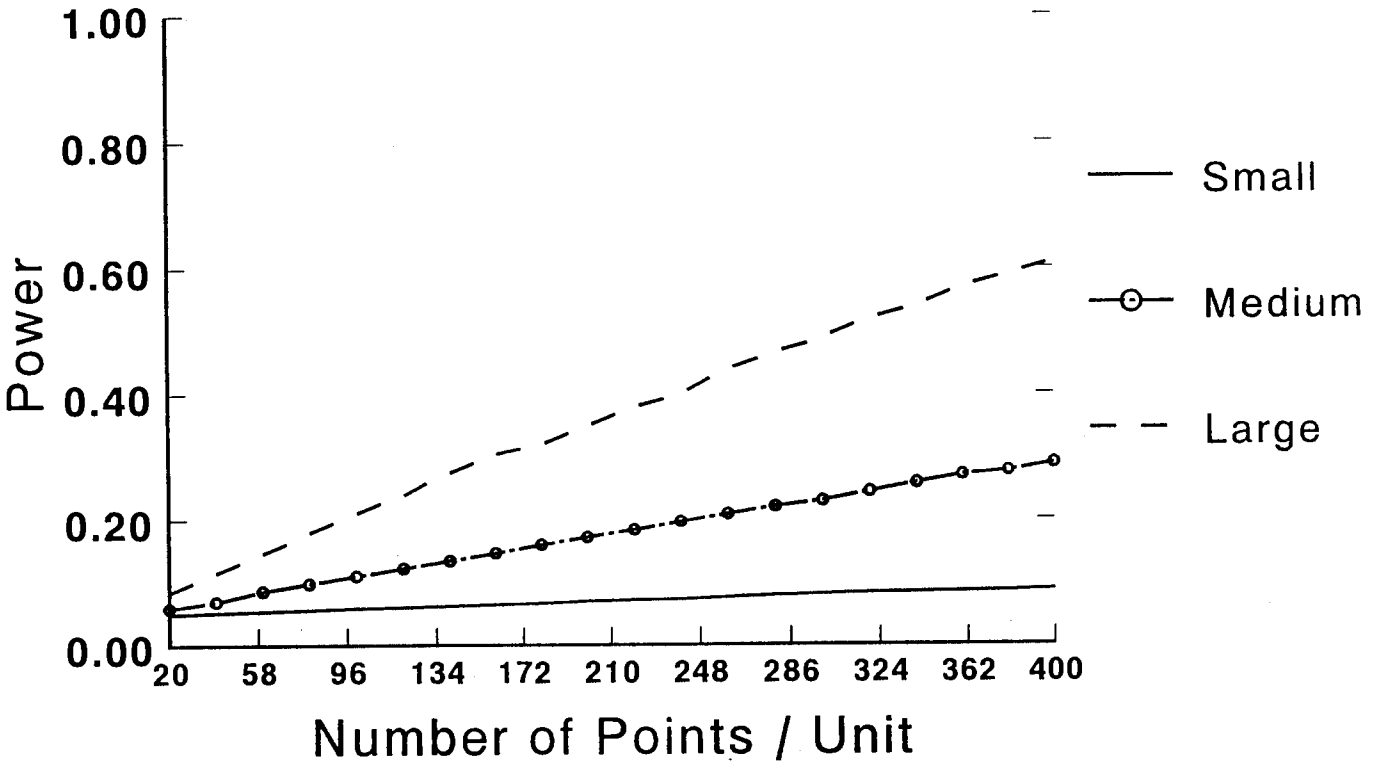
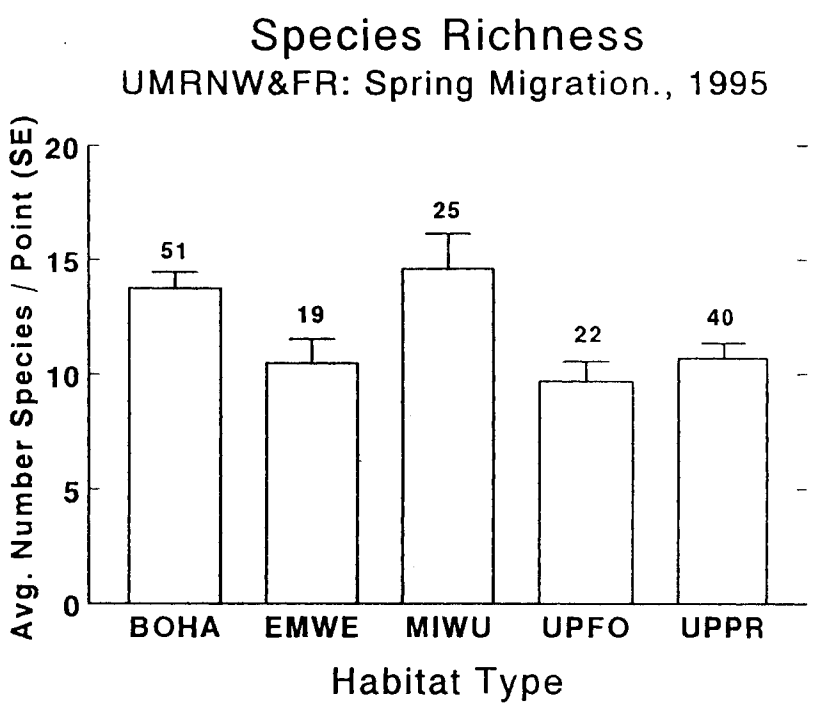
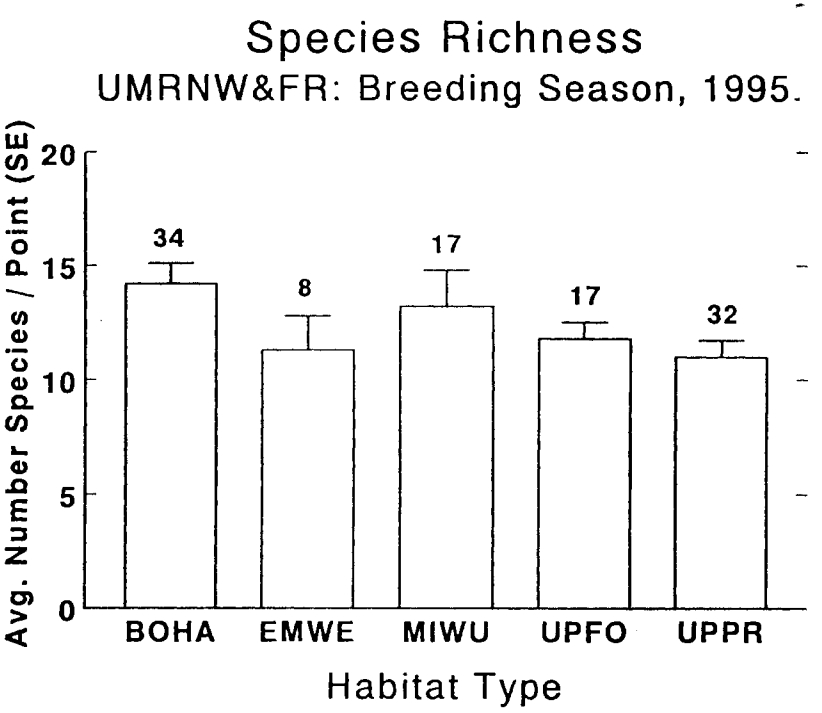


Fig. 8. Power analysis for detecting differences in overall abundances (average number of individuals/point) between refuges. Lines represent small, medium, and large "effect" sizes and roughly correspond to 20%, 50%, and 80% differences, respectively.





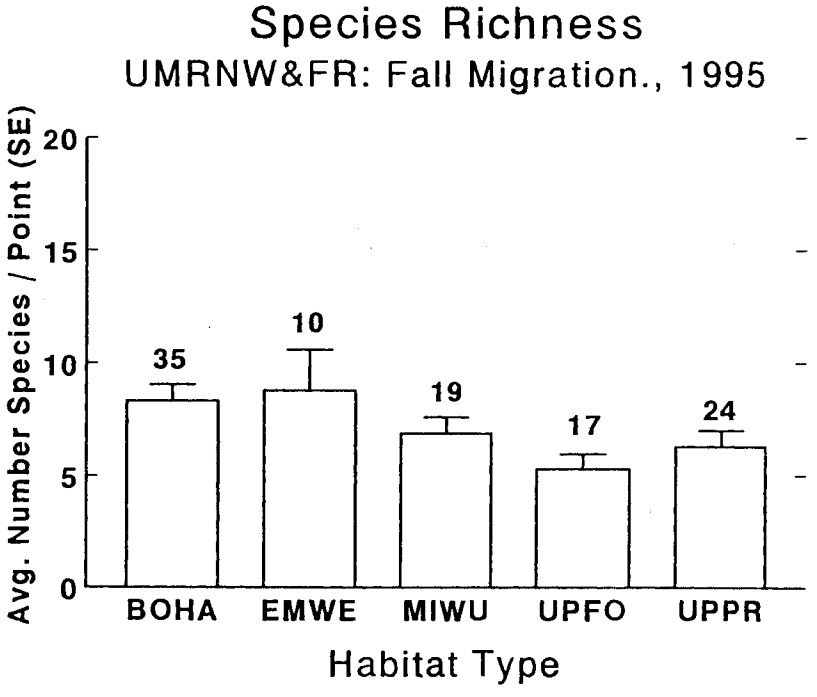


Fig. 9. Variation in species richness among habitats for 1995 census on the UMRNW and FR over three sampling periods. Sample sizes are shown on tops of bars.

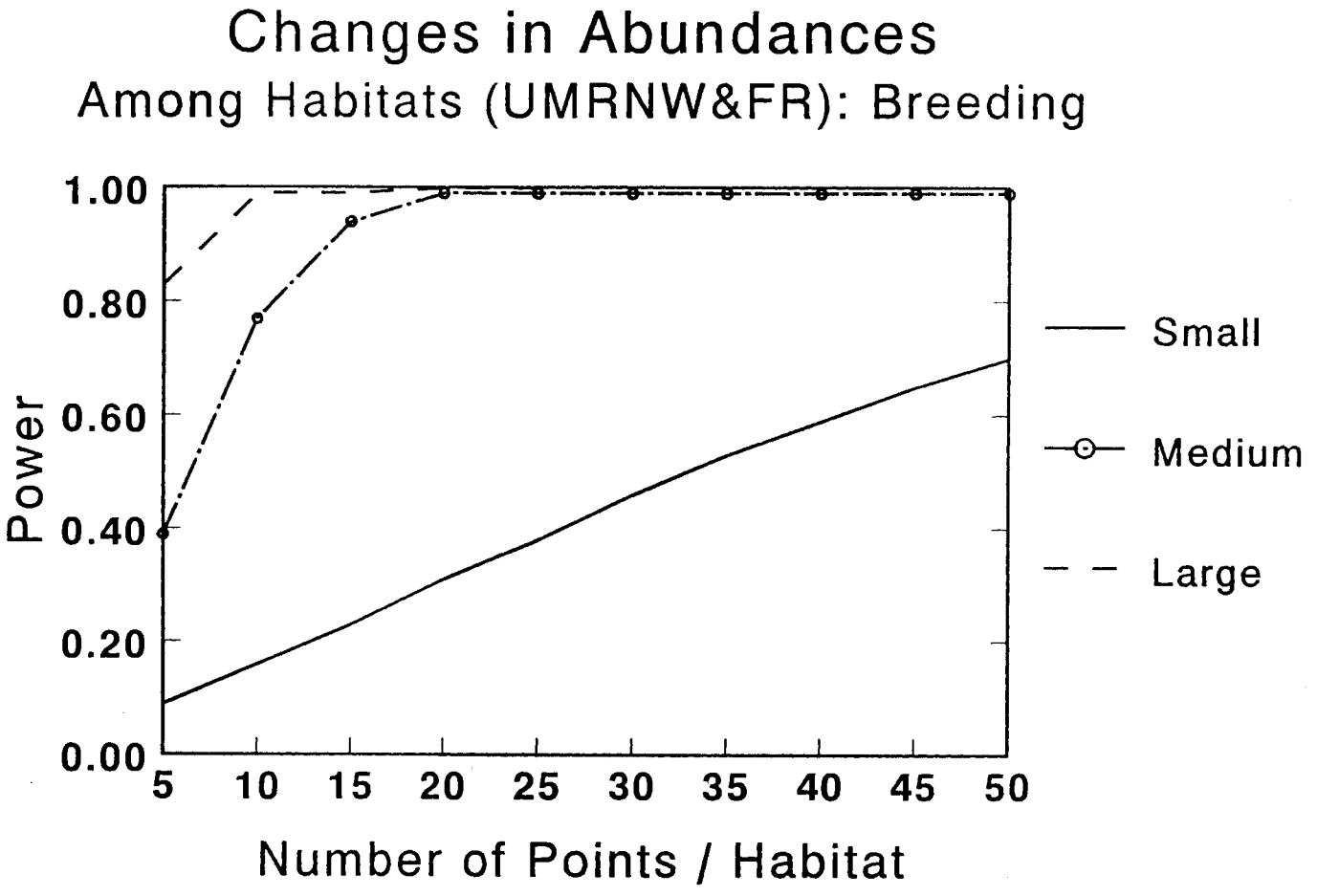
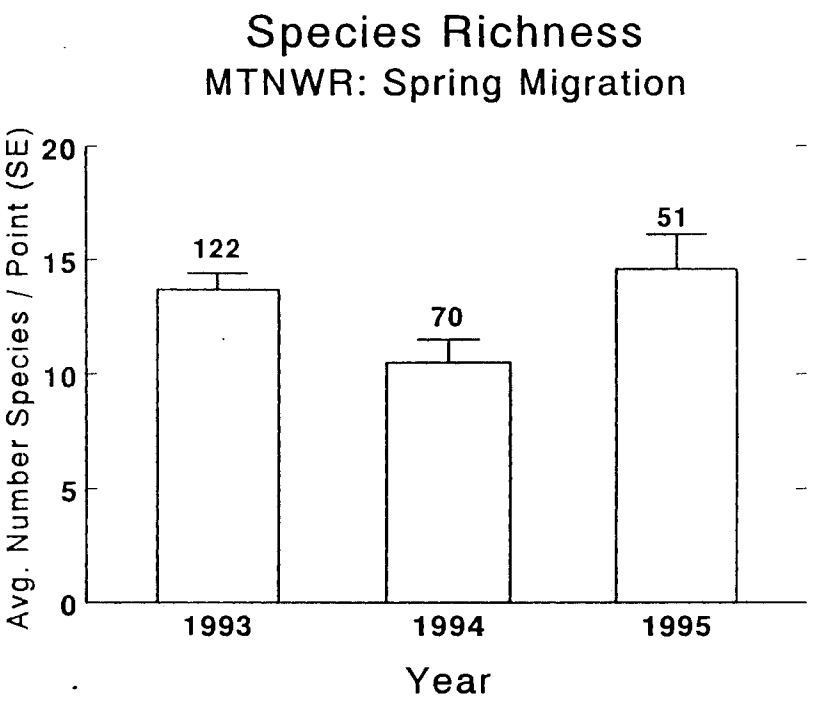
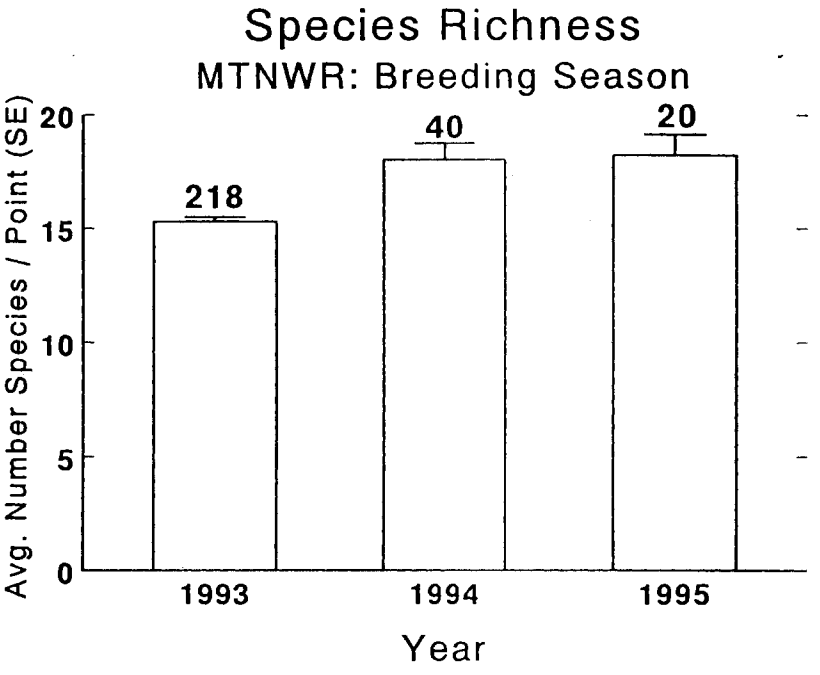


Fig. 10. Power analysis for detecting differences in species richness among habitats within the UMRNW & FR. Lines represent small, medium, and large "effect" sizes and roughly correspond to 20%, 50%, and 80% differences, respectively.

Overall abundances (again, simply the number of individuals detected / point) in 1995 were also greater on the MTNWR during all census periods (Fig. 7). These differences were significant and power to detect these differences was very high (i.e., estimated power = 1.0) for the fall and spring migration periods. For the breeding season, however, differences in overall abundances were only marginally significant (t-test, P = 0.072) and power to detect differences (at the α level of 0.05) was only 0.53. Power analyses for the two migration periods over a range of sample sizes were similar to those shown in Fig. 6. Results for the breeding season are shown in Fig. 8. and suggest that large samples would be needed to detect small and moderate differences in overall abundances between the two refuges.

To assess statistical power for comparisons of species richness or species abundances among different habitats, we considered sample data from the UMRNW&FR in 1995. Average number of species observed / point within each habitat are shown in Fig. 9. for all census periods. Habitat differences in species richness were most pronounced during the spring census period (ANOVA, $F_{4,152} = 4.98$, $P = 0.0008$), followed by the fall migration period ($F_{4,100} = 2.56$, $P = .043$) and the breeding season ($F_{4,103} = 2.12$, P = 0.084). Power to detect observed differences as significant at the 0.05 α level were 0.95, 0.82, and 0.75, respectively. Power analyses over a range of samples sizes for habitat differences in species richness during the breeding season (Fig. 10) indicate that large and medium variation among habitats would be detected with near 100 % power at samples sizes of 25 or more within each habitat. Power to detect small differences would be about 0.35 at





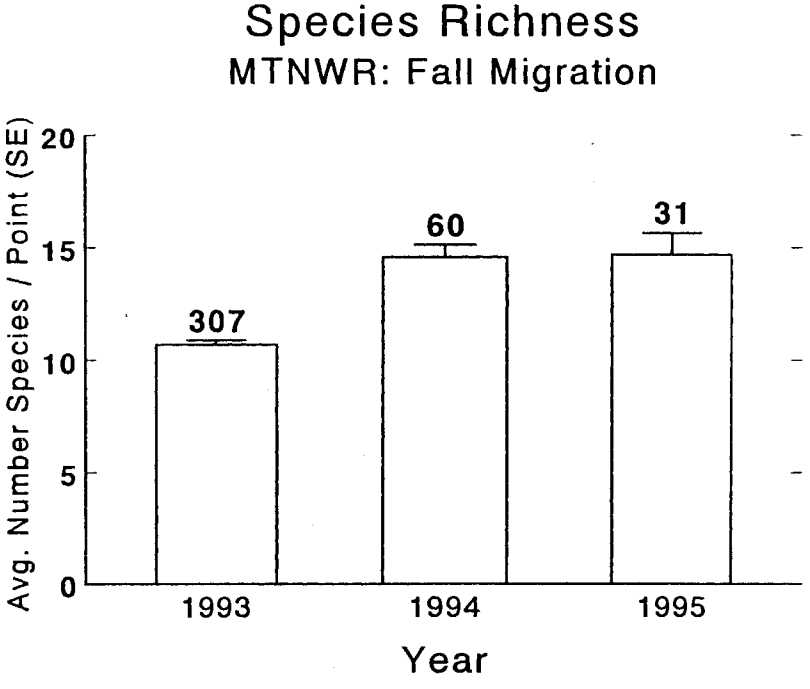
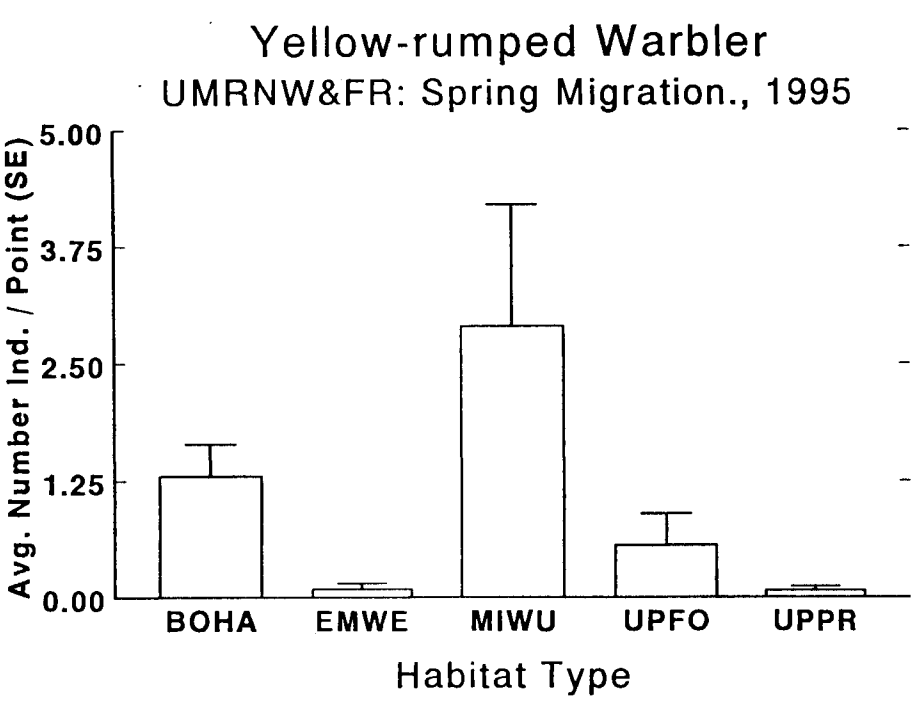
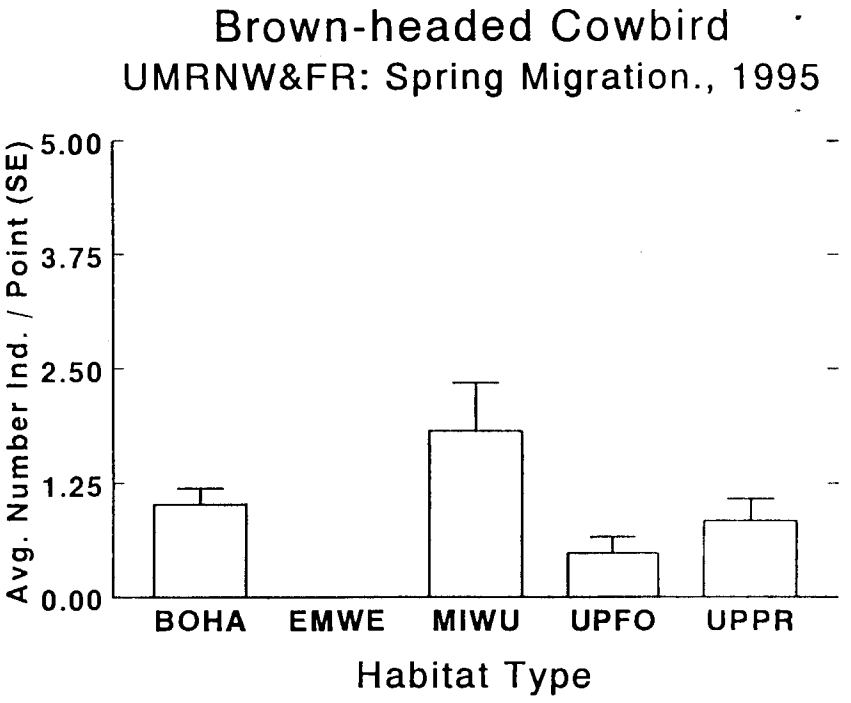
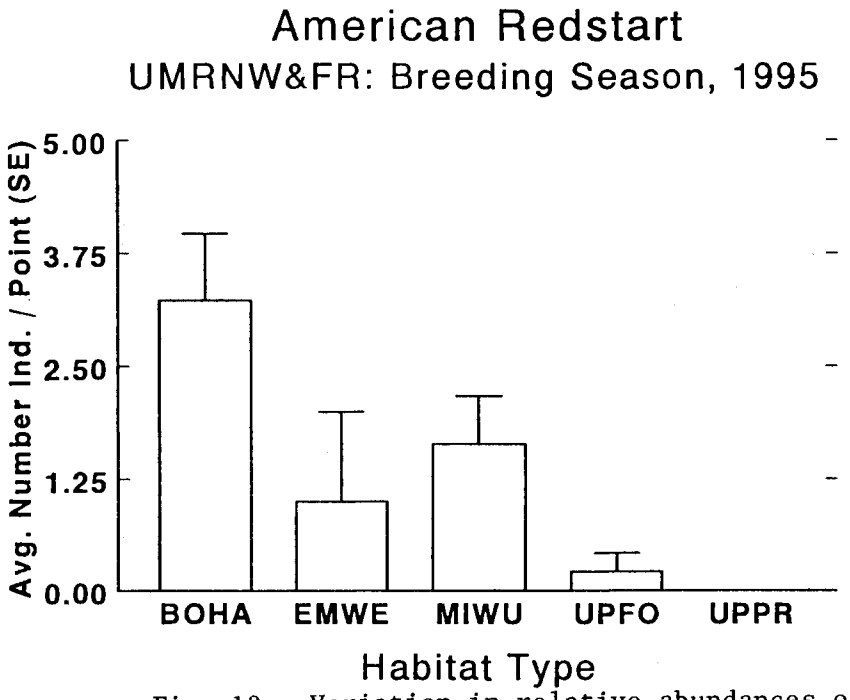


Fig. 11. Variation in species richness on the MTNWR from 1993-1995 within three sampling periods. Sample sizes are shown on tops of bars.







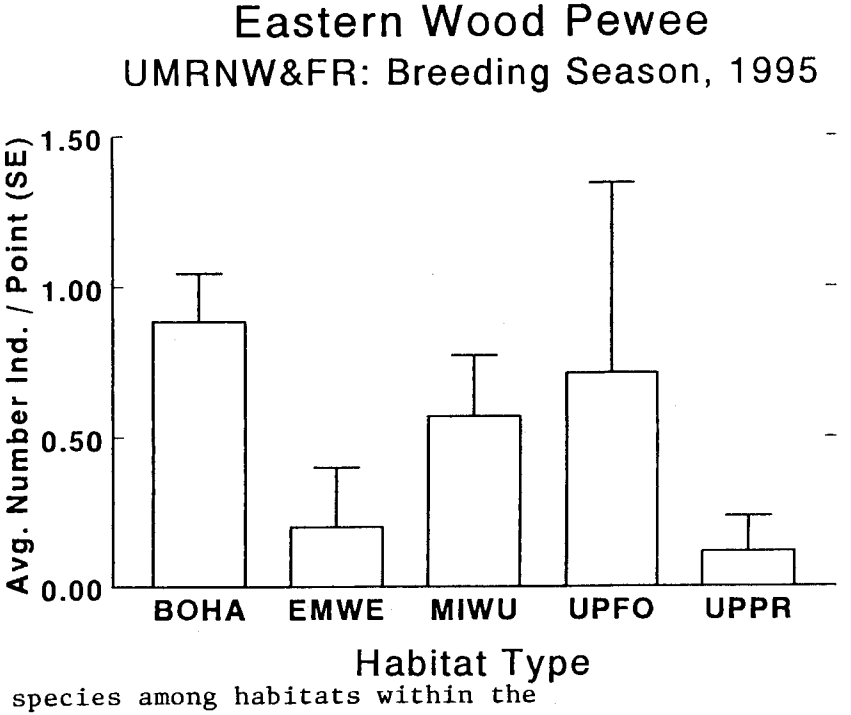
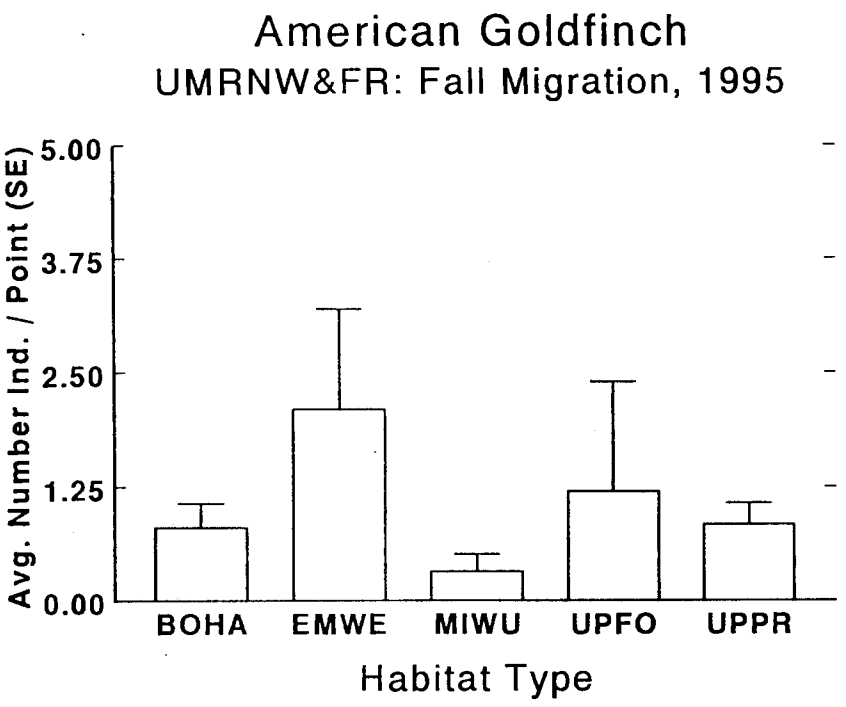


Fig. 12. Variation in relative abundances of selected species among habitats within the UMRNW & FR in 1995. Sample sizes are as shown in Fig. 9.


American Goldfinch
UMRNW&FR: Fall Migration, 1995
Avg. Number Ind. / Point (SE)
5.00
3.75
2.50
1.25
0.00
BOHA
EMWE
MIWU
UPFO
UPPR
Habitat Type

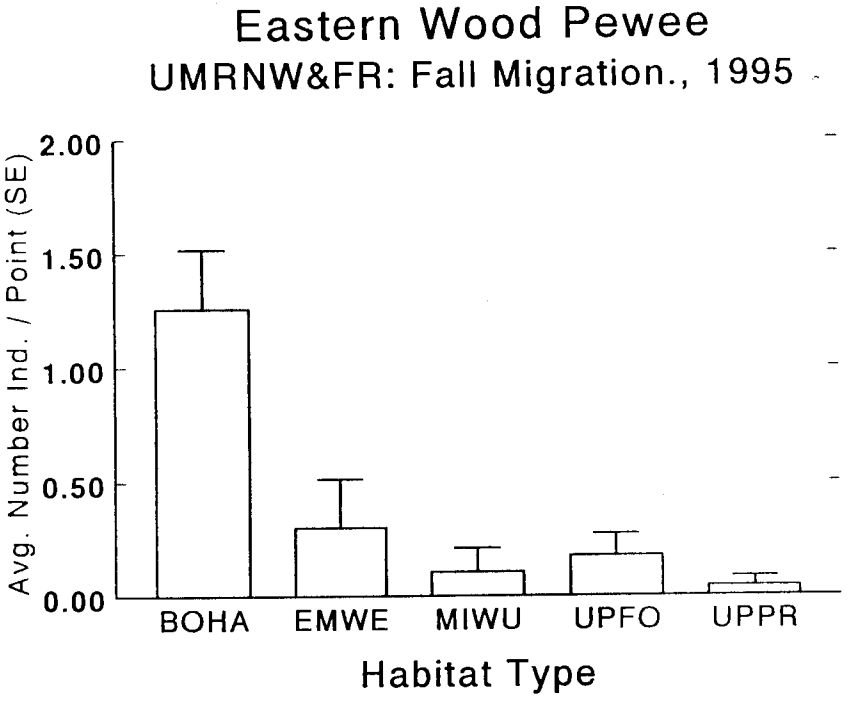

Eastern Wood Pewee
UMRNW&FR: Fall Migration., 1995
Avg. Number Ind. / Point (SE)
2.00
1.50
1.00
0.50
0.00
BOHA
EMWE
MIWU
UPFO
UPPR
Habitat Type

Fig. 12. Page 2

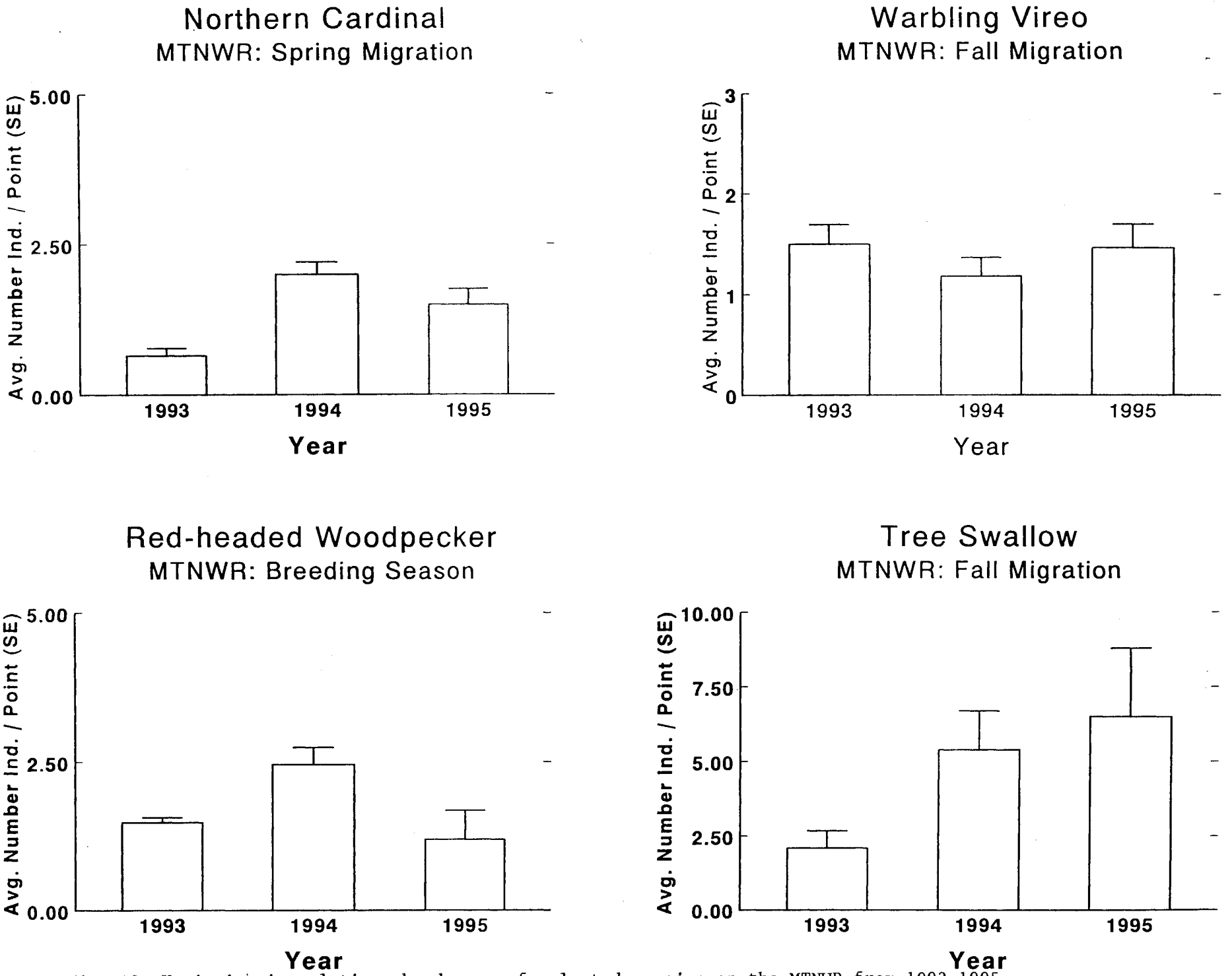


Fig. 13 Variation in relative abundances of selected species on the MTNWR from 1993-1995 within three sampling periods. Sample sizes are as shown in Fig. 11.

Table 4. Summary of power analyses for estimating variation in species abundances among habitats and years.

Season	# of Analyses	Effect sizes $\bar{X}$ (range)	Power $\bar{X}$ (range)
Spring	10	.52 (.22-.68)	.78 (.35-.99)
Breeding	10	.63 (.29-.89)	.84 (.32-.99)
Fall	10	.99 (.36-.74)	.74 (.20-.98)

sample sizes of 25 and about .75 for habitat samples sizes at large as 50. Note that these power estimates were derived assuming unequal sample sizes among habitats.

We assessed power to detect variation in species richness over time with data from the MTNWR (1993-1995), and one-way ANOVA. We did not carry out regression-type trend analyses because too few years are available at this time (options for analyses of temporal trends in species richness or abundances are discussed below). Species richness varied significantly by year for all sampling periods (Fig. 11, F-tests, $P < 0.01$). Power to detect these differences was near 1.0. Analyses of a range of sample sizes for a period covering five years indicated that, for each sampling period, samples of 100 or more would yield power of near 1.0 for even small effects.

We approached power analyses for changes in species abundances by selecting certain species and, as above, considering variation over habitat and time. Selected patterns of variation among habitats for the three sampling patterns are illustrated in Fig. 12. Not surprisingly, nearly all the habitat comparisons we selected revealed significant differences in abundances among at least two of the habitat-types (Bonferroni tests). Within-habitat sample sizes for these tests ranged from about 10 to 60. Power to detect observed differences as significant at the 0.05 level was generally above 0.70 and in most cases was greater than 0.90 for all sampling seasons (Table 4). Selected examples of annual variation within the MTNWR (Fig. 13) also indicated that expended sampling effort was sufficient to detect moderate and large effects. For small effects, power was still generally above 0.50.

Another approach to power analyses is one developed by Gibbs (1995) where a Monte Carlo approach is used to estimate power in detecting trends in survey data such as

those taken in the UMR. At present, too few years of data are available to analyze existing trends over time. Nonetheless, we used the data to provide some of the necessary parameters (i.e., "initial values") and estimated relationships between sample size and power. We considered each sampling period separately, selected five species, and calculated power to detect linear trends (up or down) with sample sizes of 50, 100, 150, and 200 points. For these analyses, we assumed single visits to points within each season . Results of these simulations revealed that with sample sizes of 100 or more, power to detect (linear) trends of 4% annual change or more (up or down) was uniformly above 0.70 and typically > 0.80.

In summary, the sample sizes expended in the UMR monitoring program appear to be sufficient with respect to detection of species and characterizing species composition, precision of estimates (species richness or abundances), and power to detect spatial and temporal variation in either species richness or specific abundances. Options for analyzing differences among habitat, among years, and possible interactions between the factors are discussed below

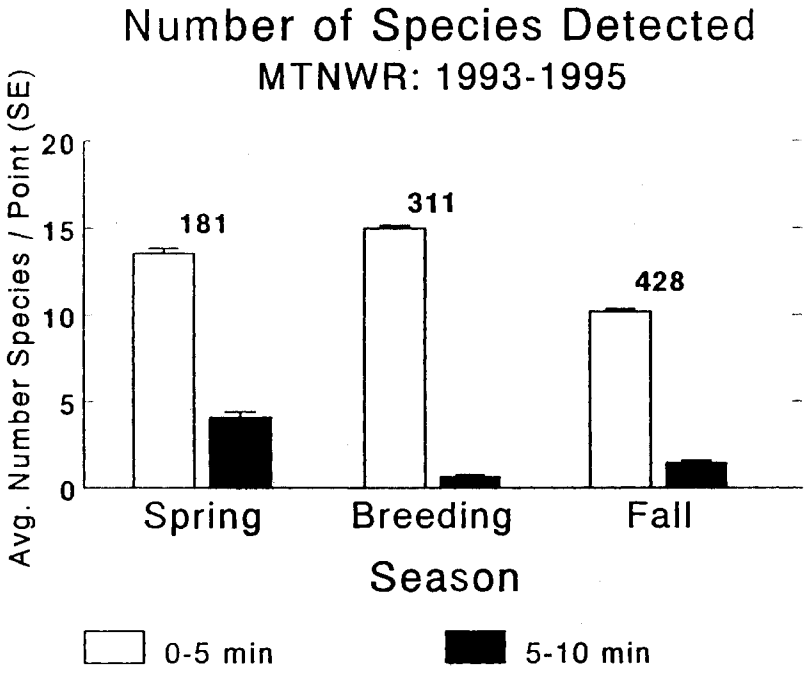
Analyses of point count methodology

Two questions arise in any point count program with respect to methodology: how long and how far? For the UMR monitoring program, the important questions are 50 versus 100 m radius counts and 5 versus 10 minute counts. To evaluate sampling at different distances in the UMR program, we compared numbers of species and individuals observed at 0-50 m and at 50-100 m. For this analysis, various subtotals in the database (i.e., 0-25 an 25-50) were summed. We performed paired t-tests to assess if significantly different

numbers of species or individuals are detected within these to distance bands and whether the 0-50 m subtotals are significantly different from the overall totals ("outside" observation were not counted for these analyses). For the species counts, we totaled species that were observed uniquely in each band; thus, the subtotal for 50-100 m was species that were added by sampling out to that distance. For these analyses, we used data from the MTNWR and present only analyses for the 5 minute subtotal (about 80-90% of the detections were within the first 5 minutes and analyses of distance using the total 10 minute were nearly identical).

Not surprisingly, significantly more species and individuals were observed from 0-50 m band than from 50-100 m (Fig. 14, paired t-tests, $P < 0.05$). For both variables the difference was least pronounced within the fall sampling period (Fig. 14). For all sampling periods, total counts for both species and individuals were significantly greater than the 0-50 m subtotal; therefore, the extra distance had a significant effect. Notwithstanding, one factor underlying this result is the large sample sizes (up to 450) used for the paired t-tests. Even very small mean differences will be judged significant with large sample sizes because standard errors that accompany large sample sizes are typically small.

Comparisons of 5 versus 10 minute counts were carried on data from the UMRNW&FR. Analyses were conducted as above and revealed that the longer counts had more species and individuals detected / point (Fig. 15) For species and individuals, the relative contribution of the second 5 minutes was greatest with the spring counts and least during the breeding season. In all cases, the 5-10 minute period significantly increased total number of species or numbers of species detected (paired t-tests, $P < 0.05$).



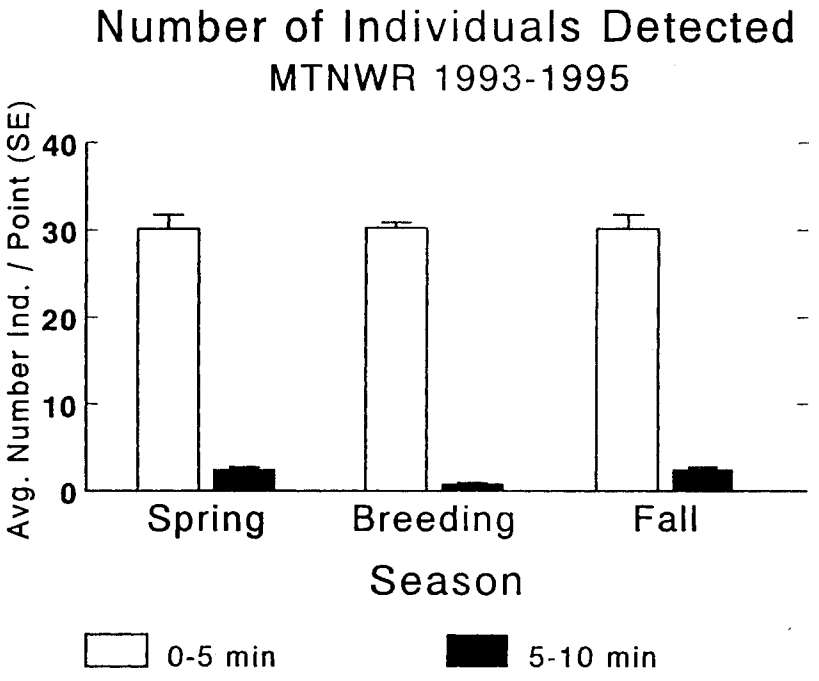


Fig. 15. Comparisons of numbers of species and individuals detected during two time intervals on the MTNWR. Sample sizes are shown in graph on the left.

Recommendations

Sample Size and Selection of Sampling Points. Recommendations for sample size are made on a habitat and refuge wide basis, and are based on rarefaction, bootstrapping, and power analyses. Habitat recommendations are for within refuges. For the rarefaction analyses, we estimated where, on a given curve, the expected number of species did not appreciably increase with more sampling. We then took the number of individuals sampled at that point and divided by the average number of individuals observed / point to derive a suggestion for the number of points needed to adequately characterize species composition.

Recommendations for minimum sample sizes are offered in Table 5. Note that the recommended sample sizes vary among seasons and habitats. Generally, sample sizes need to be greater in areas where species richness is higher, where - on average - fewer individuals are observed / point, or where variation among points is greater. Sampling needs within each habitat sum to the total recommended for district to district comparisons. We recommend that the refuge-wide number of samples be adopted along with a protocol for stratifying sampling effort by habitat (see below). The interim guidelines for the UMR monitoring program call for 60 points / habitat type; the recommendations in Table 5 are in general agreement with this guideline.

The process by which points are selected in a monitoring program or any sample is critical. As recommended in the current guidelines for the UMR, some element of randomization is a general principle to adhere to. Without randomization, the biological

Table 5. Recommended minimum sample sizes (# of points) for Upper Mississippi monitoring program. All recommendations are for number points/refuge where habitat is available.

Area or Habitat	Season-Spring	Season-Breeding	Season-Fall
Upland Forest	40	50	60
Bottomland Hardwood	60	50	60
Upland Prairie	70	40	50
Emergent Wetland	40	40	60
Mixed Upland/Wetland	40	35	50
Refuge Wide	250	225	280

importance of inferences from statistical analyses are questionable at best. Points selected without randomization often reflect judgments by observers that do not lend themselves to unbiased statistical results. For birds, it is well known that "birders" will tend to select sites where bird diversity or abundances are relatively high (S. Robinson, *personal communication*). Ideally, biological expertise and randomization play key roles in selection of sampling material.

Estimating trends over time and among habitats are two major objectives of the UMR monitoring program. Moreover, unbiased estimates of trends within habitats and over entire districts (or refuges) are needed. The need for habitat-specific data calls for a stratification procedure as recommended in the interim guidelines; otherwise a systematic sample can be satisfactory (Ralph 1995). The stratified (on habitat) randomization process recommended in the interim guidelines for the UMR has much to recommend it. Nonetheless, with volunteer efforts or limited resources, logistics play an important role. Travel time among points can be a limiting factor for purely random surveys (Pendleton 1995) - especially in some areas within the UMR observers cannot easily walk or drive from point to point. A complicating factor for the UMR is that certain habitats or districts will easily accommodate a stratified randomization procedure while others will not. Thus a single prescription may not be widely applicable.

With these limitations acknowledged, we offer the following guidelines for selection of points. Note that - where possible - we adhered to the existing interim guidelines.

I) Stratified random sample where access and travel times are not limiting:

1) Identify habitats of biological/management importance. At present, five habitats have been identified. These designations can be modified, but should be relevant throughout the entire UMR and agreed upon by appropriate personnel. Different habitat designations for different refuges or districts are inadvisable.

2) Identify above habitats within each refuge or appropriate management unit and identify a population of points that can be sampled. Selection of sampling points from these populations (without replacement) using numbered grid blocks of homogeneous habitat with 250 m (minimum) radii as outlined in the interim guidelines are a suitable method. The randomization methods in the interim guidelines are also acceptable. Importantly, the selection process within each stratum should be independent of other strata.

3) Two options for allocation of points within strata are fixed-sample size or proportional allocation. With fixed-sample size, sample sizes at least as large as those recommended in Table 5 should be assigned to each strata (in cases where the minimum varies among sampling seasons, use the greatest number). The disadvantage of this design is that some habitats might tend to be over or under sampled. With proportional allocation, the fraction of the overall sample size within each stratum is proportional to its representation in area. Therefore, if bottomland hardwoods comprise 40% of the holdings within a refuge, 40% of the points should fall within that habitat. A total sample size should be decided upon beforehand (proceed

with the assumption that each point will be visited only once within a season). A possible disadvantage for this method is that minimum sample sizes for a given habitat may not be achieved. Adding a few points to these habitats would be acceptable.

4) The same points can be used for sampling in the three sampling seasons.

5) After census points are selected, design a route and sampling schedule.

With the above procedure, sample means and variances within each habitat can be combined (by assigning weights to each stratum) to yield refuge-wide and unbiased parameter estimates (Krebs 1989). If sampling within strata is truly random, then the overall sample will be representative of the entire refuge (Thompson 1992).

II) Stratified random sample where access and travel times are limiting:

We strongly recommend that the sampling process use the above procedures; however, if logistics pose a unavoidable constraint, we recommend the following adjustments.

1) Use on-road counts. The above protocol is for off-road counts. If secondary and tertiary road are available, then on-road counts can be carried out. Previous studies (e.g. Buskirk and McDonald 1995) indicate that counts from small roads and on-road counts yield similar estimates of abundance and species richness. Selection of points on the road can be randomized by selecting points as above and going to the nearest adjacent road. This method is not advisable if only certain habitats are accessible by roads.

2) Rotate through subsets of points over years. With this adjustment only subsets of points would be visited each year. The number of points visited each year would be dictated by logistics. With this protocol, spatial and temporal trends could still be assessed, but the procedures would need to be modified.

3) **Use transects of points where the first point in the transect is randomly selected**. With this protocol, transects of point (e.g., 10 points) with at least 250 m intervals would be established within each habitat. The starting point for each transect would be chosen randomly. The direction of the transect will need to be determined so that homogeneous habitat is censused. This procedure would reduce travel time among points.

Sampling procedures at each point. Sampling efficiency at each point can greatly influence the success of a monitoring program. The presently used protocol for point counts in the UMR program is basically sound. Nonetheless we do recommend changes that we believe will lead to more information / sampling unit.

Timing: Analyses presented above and other studies (Ralph et al. 1995) indicate that 10 minutes are appropriate. If the number of points that can be visited is appreciably less than that recommended above, then longer counts of 15 or 20 minutes may be necessary. A disadvantage of counts longer than 5 minutes is that birds may be counted more than once owing to movements during the count period. This adjustment to longer counts will be especially important for characterizing species composition. Regardless of the duration, we recommend that subtotals for three and five minutes be recorded and

retrievable from the databases. These subtotals will maximize the usefulness of the UMR data for comparisons with data from other programs such as the Breeding Bird Census.

Distance: At present, birds observed out to 100 m are recorded and the databases include 50 m subtotals. We recommend unlimited-distance point counts and also recommend that the direction of the bird from the observer be recorded (direction can be easily recorded with a compass on the clipboard). We recognize these changes will require highly-trained observers, but we recommend these changes for the following reasons. First, by limiting the count to 100 m, many individuals are not counted except as "outside." Totals in the outside category were often high and we believe that the efficiency of counts will rise greatly with unlimited-radius counts. Not counting birds that are visible and heard at, say, 125 m sacrifices much information. Second, data from unlimited radius counts can be used or converted to fixed-radius counts if the distance from the observer the bird is estimated and recorded in the database. A simple filter to use only those birds counted within 100 m or 50 m can be easily applied. With unlimited-radius counts, birds that are so far away that a distance cannot be reasonably estimated are still counted as "outside."

Third, an assumption of point counts (and most other counts based on detection of singing birds) is that the probability of detection for different species or different "types" of individuals within a species (e.g., mated versus unmated males) is equal. Therefore, a disadvantage of point counts is that the indices derived may have some important biases. Seasonal changes in singing rates within a species are also ignored. The assumption of "equal detectability" is almost certainly false; certain species sing more often or louder than others, mated males sing less often, and singing rates typically decrease throughout

the breeding season (Verner 1985). Therefore, serious biases are inherent in an indexed estimate of abundance such as point counts. The problem is that the sampling process is not modeled and no attempt is made to adjust for differences in detectability; without modeling, the census efforts produce an "unadjusted count." The problem of heterogeneity in detectability, is the basis for the Jolly-Seber approach to demographic analysis, and attempts to rectify the problem are what led to the development of other methods such as the variable circular-plot and the Emlen line-transect method for censusing birds (Verner 1985). If detection distances and directions are recorded , the samples can be analyzed as point counts and, if desirable, variable circular plots. The latter method carries the advantage of having the ability to model the sampling process and produce estimates corrected for differential detection probabilities.

Number of visits / point / season. - We follow Ralph et al. (1995) in recommending that each point be visited only once / season. Single visits will allow more sites to be visited and increase coverage within the UMR.

Suggested Analyses

Trends in Species Abundances or Richness. Depending on the biological question, point count data support several methods of analysis. Discussions with UMR personnel indicate that the primary questions where statistical analyses would be needed are to assess differences in species abundances or species richness among habitats and trends in the quantities over time. The latter could apply to specific habitats, refuges or the entire UMR.

For a single year, variation among habitats, refuges, districts, etc. could be assessed by the usual ANOVA-type approach. Variations in the standard one-way ANOVA routine

that accommodate unbalanced designs or unequal variances are available in most statistical software packages. If sample distributions deviate seriously from assumptions of normality (which could arise easily with rare species and "0" counts at many point), then nonparametric analogues could be used - typically with slight loss in efficiency and power.

Analyses of habitat or spatial variation over several years will require a different model of analysis. A simple two-way ANOVA with, for example, habitats and years as the factors would be inappropriate if the same sampling point are visited over time. As we are not recommending that a new set of points be selected each year, repeat visits are likely to be the case. Therefore we recommend repeated-measures ANOVA (r-m ANOVA). With this model it is possible to assess variation in abundances among "groups" of sampling plots (i.e., between subjects) and over time (i.e., within subjects). Groups of sampling plots could correspond habitat types, migratory status, etc. Variation over time would correspond to annual variation. An advantage of this model is that biologically interesting interactions between spatial and temporal variation can be assessed. Details on this type of model can be found in Milliken and Johnson (1984). Most questions that arise in monitoring programs can be evaluated with this design, including pairwise comparisons of selected years or strata of points using linear contrasts. Importantly, r-m ANOVA will identify significant variation through time. Such variation does not necessarily translate into a "trend," however. For example, if a sample covers 6 year and abundances rise the first 3 years, but decrease the next 3 years, a significant time effect will likely be detected. A r-m ANOVA will also permit analyses of biologically interesting time x group interactions whereby - if present - trends through time may follow different patterns for different habitats.

Another option for simple trend analysis is a regression approach as presented by Gerrodote (1987, 1991). This approach estimates % changes over time (or over space) and tests for the significance of the estimated trend. At present, too few years of sampling have been completed to use this method effectively. An advantage to this option is that a software package called TRENDS is available to perform the calculations (T. Gerrodette, Southwest Fisheries Science Center, P.O. Box 271, La Jolla, CA 92038). An uncertainty to his approach is whether use of the same point over time leads to pseudoreplication, correlated errors, and inflated p-values.

Many nonparametric techniques for the analyses of trends over time have been developed and these circumvent some of the assumptions that may be difficult to achieve with parametric procedures. Many of these techniques are based on the Mann-Kendall trend statistic and are computationally relatively simple. Berryman et al. (1988) review these techniques and offer guideline for their use.

One caveat mentioned above, but very relevant to a discussion of analytic options is the problem of changes in detectability. If detectability changes over time owing to changes in personnel or changes in observer skills, then serious biases are introduced and inferential statistics, by any approach must be interpreted with caution (Barker and Sauer 1995). Changes in detectability can also occur among habitats. This problem (which is not unique to point counts) suggest that exploratory analyses may also be appropriate for analyses of the UMR data. Visual displays that convey information about patterns in abundances and specie richness may be as effective and meaningful as inferential statistics in guiding management policies.

Bird-habitat associations. We did not perform multivariate analyses to establish associations between avian abundances and plants or habitat structure. For the UMR protocol as it now exists, we urge caution for analysis of bird-habitat associations. Locations of signing males may simply indicate the type of habitat where unmated birds sing throughout the breeding season. This habitat may differ from that where birds nest or where territories of actual breeding pairs are located. Although the "number of pairs" is the stated metric, the relationship between the actual number of breeding pairs and the count indices collected in the UMR program is unknown.

Notwithstanding the above problem, we recommend an exploratory/experimental approach to analyses of bird-habitat relationships. Multivariate ordination procedures such as principle components, or detrended correspondence analysis are appropriate for identifying the specific variables that underlie variation in habitat. These variables can then be related to avian abundances by multivariate regression for abundances or species richness and logistic regression for simple presence absence. We believe that results of these analyses should be used for exploratory purposes and serve as the basis confirmatory field studies. When feasible, habitat variables or suites of variables identified as important sources of variation in avian abundances should be candidates for management and manipulation. With forethought, such manipulations can be accomplished by utilizing habitat alterations (before and after) associated with management or disturbances and insuring that habitat on non-impacted areas or plots are measured as well. Without such studies, we recommend against the use of multivariate habitat associations as the sole basis of management policies/recommendations on the UMR.

The UMR sampling as it now exists does present an excellent opportunity to assess covariation through time in avian abundances (or nesting success, see below) and various local habitat variables. These temporal changes in habitat can be associated with succession or be directly man-induced. Techniques such as cross-correlation analysis will evaluate relationships between two time series (with different lag periods). Simple partial correlation analyses among suites of variables are useful in establishing patterns of covariation among variables, but hypothesis testing is not advised owing to lack of independence (i.e., autocorrelation) among years.

Recommendations for Data Management

We recommend that the number of fields in the bird databases be increased or - alternatively - relations and utilities be created to extract more information from the census data than is now possible. An important enhancement is to associate species with a suite of ecological, taxonomic, and other life history traits. A suggested list of these traits is offered in Table 6. With these attributes, biologically interesting comparisons can be made. For example, it may be of interest to assess whether the patterns of change through time are different for neotropical migrants than for short distance migrants or permanent residents. Another possibility is to compare groups of species with different generalized habitat associations (e.g., grassland versus forest versus open woodland) or different foraging ecologies.

We also strongly recommend that observer identity be added to all of the bird databases.

Table 6. Suggested additions for UMR monitoring databases. Items would be added to each record

Observer
Migratory Status
Foraging Guild
Nesting Guild
Body Size
Family
Conservation Status (Regional Partners-in -Flight "Priority Scores" could be used)

Recommendations Enhancements for Field Methods

Our major recommendation for UMR monitoring of birds is that program be integrated with other programs with compatible objectives. For mobile organisms like birds, local changes in avian abundances may not stem from local changes in productivity or availability of habitat; rather, local trends may reflect regional patterns and dynamics (Brawn and Robinson 1996). For example, a steep decrease in the abundances of several species of neotropical migrants on an installation might be part of a regional decrease owing to changes in habitat on wintering grounds in the Neotropics. Therefore, monitoring of reproductive success and vital demographic parameters such as survival rate and annual recruitment is needed. The programs known as BBIRD and MAPS monitor these quantities. Serious consideration should be given to establishing a series of MAPS and BBIRD sites in the UMR corridor.

Another need that is somewhat unique to the UMR is more detailed information on the use of corridors during the migration periods. At present, abundances are monitored, but specific use of the corridor is not. Information on foraging during stopover with respect to tree-species use would be invaluable. In comparison with the habitat ecology of NMFBs during the breeding season, habitat needs during migration are poorly understood.

ACKNOWLEDGMENTS

We thank the following personnel for their cooperation, time, and advice on this evaluation: Eric Nelson, Jim Mattsson, Melinda Knutson, Jim Quinlivan, Eileen Kirsch, Bea Treiterer, and Steve Lewis. We also thank B. Gourley for assistance in preparation of the report.

LITERATURE CITED

Barker, R. J. and J. R. Sauer. 1995. Statistical aspects of point count sampling. Pages 125-130 *in* C. J. Ralph, J. R. Sauer, and S. Droege (technical editors). Monitoring bird populations by point counts. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-149. 187p.

Bibby, C. J., N. D. Burgess, and D. A. Hill. 1992. Bird census techniques. Academic Press, New York.

Berryman, D., B. Bobee, D. Cluis, and J. Haemmerli. 1988. Nonparametric tests for trend detection in water quality time series. Water Resources Bulletin 24:545-556.

Brawn, J. D., and S. K. Robinson. 1996. Source-sink population dynamics may complicate the interpretation of long-term census data. Ecology 77:3-12.

Buskirk, W. H., and J. L. McDonald. 1995. Comparison of point count sampling regimes for monitoring forest birds. Pages 25-34 *in* C. J. Ralph, J. R. Sauer, and S. Droege (technical editors). Monitoring bird populations by point counts. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-149. 187p.

Butcher, G. S., B. Peterjohn, and C. J. Ralph. 1993. Overview of national bird monitoring programs and databases. Pages 192-203 *in* D. M. Finch and P. W. Stangel (eds.). Status and management of neotropical migrant birds. USDA For. Serv. Gen. Tech. Rep. RM-229.

Efron B. & R. Tibshirani. 1993. Introduction to the bootstrap. Chapman and Hall.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364-1372

Gerrodette, T. 1991. Models of power for detecting trends - a reply to Link and Hatfield. Ecology 72:1889-1872.

Gibbs, J. 1995. Monitor Users Manual (available for the author: JamesGibbs@aol.com)

Gotelli, N. J. and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington.

Hutto, R. L., S. M. Pletschet, & P. Hendricks. 1986. A fixed-radius point count method for nonbreeding and breeding season use. Auk 103:593-602.

Krebs, C. J. 1989. Ecological methodology. Harper Collins, New York.

Lancia, R. A., J. D. Nichols, and K. H. Pollock. Estimating the number of animals in wildlife populations. Pages 215-253 *in* T. A. Bookhout (ed.). Research and management techniques for wildlife and habitats, 5th ed. The Wildlife Society, Bethesda, Maryland.

Ludwig, J. A. and Reynolds, J. F. 1988. Statistical ecology: a primer on methods and computing. John Wiley and Sons, New York.

Millikin, G. A. and D. F. Johnson. Analyses of messy data, volume 1. Van Nostrand Reinhold. New York.

Pendleton, G. W. 1995. Effects of sampling strategy, detection probability, and independence of counts on the use of point counts. Pages 131-135 *in* C. J. Ralph, J. R. Sauer, and S. Droege (technical editors). Monitoring bird populations by point counts. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-149. 187p.

Ralph, C. J. G. R. Geupel, P. Pyle, and T. E. Martin. 1993. Handbook of field methods for monitoring landbirds. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-144. 41 p

Ralph, C. J., J. R. Sauer, & S. Droege. 1995. Monitoring bird populations by point counts. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-149. 187 p.

Thompson, S. K. 1992. Sampling. John Wiley and Sons.

Verner, J. 1985. Assessment of counting techniques. Curr. Ornithol. 2:247-302.