

ESTIMATING SPECIES RICHNESS: THE IMPORTANCE OF HETEROGENEITY IN SPECIES DETECTABILITY

THIERRY BOULINIER,^{1,4} JAMES D. NICHOLS,² JOHN R. SAUER,² JAMES E. HINES,² AND K. H. POLLOCK³

¹*North Carolina Cooperative Fish and Wildlife Research Unit, North Carolina State University, Raleigh, North Carolina 27695 USA*

²*U.S. Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA*

³*Institute of Statistics, North Carolina State University, Box 8203, Raleigh, North Carolina 27695-8203 USA*

Abstract. Estimating species richness (i.e., the actual number of species present in a given area) is a basic objective of many field studies carried out in community ecology and is also of crucial concern when dealing with the conservation and management of biodiversity. In most studies, the total number of species recorded in an area at a given time is taken as a measure of species richness. Here we use a capture–recapture approach to species richness estimation with North American Breeding Bird Survey (BBS) data in order to estimate species detectability and thus gain insight about its importance. In particular, competing models making different assumptions about species detectability are available. We carried out analyses on all survey routes of four states, Arizona, Maryland, North Dakota, and Wisconsin, in two years, 1970 and 1990. These states were chosen to provide contrasting habitats, bird species composition, and survey quality. We investigated the effect of state, year, and observer ability on the proportions of different models selected, and on estimates of detectability and species richness. Our results indicate that model M_{jh} , which assumes heterogeneous detection probability among species, is frequently appropriate for estimating species richness from BBS data. Species detectability varied among states and was higher for the more skilled observers. These results emphasize the need to take into account potential heterogeneities in detectability among species in studies of factors affecting species richness.

Key words: capture–recapture; community ecology; detectability; heterogeneity; model selection; North American Breeding Bird Survey; species richness.

INTRODUCTION

Estimating species richness (i.e., the actual number of species present in a given area) is the basic step of many field studies carried out in community ecology and is also of crucial concern when dealing with the conservation and management of biodiversity (May 1988, Colwell and Coddington 1994). In most studies, count data (e.g., the total number of species recorded in an area at a given time) are used as an estimate of species richness (e.g., in recent bird community studies: Enoksson et al. 1995, Knick and Rotenberry 1995, McIntyre 1995, Schiek et al. 1995, Riffell et al. 1996), but see Karr et al. (1990), Dawson et al. (1995), and Thiollay (1995) for alternatives. Such studies investigate spatial or temporal trends in species richness or the effects of different environmental factors on the local occurrence of species. However to test any hypotheses about changes in species richness using data from counts of species, it must be assumed either that all species are detected (which is not true in most biological samples), or that detectability of the different species is the same or at least does not differ among

groups to be compared. The detectability of a species can be defined as the probability of detecting at least one individual of a given species in a particular sampling effort, given that individuals of that species are present in the area of interest during the sampling session.

Unfortunately, the assumption of equal detectability among species is generally false and unequal species detection probabilities can invalidate the results of hypothesis tests. A large body of methods exists to estimate species richness taking into account species that are not actually recorded but whose presence can be inferred from the pattern of observed species occurrence (Burnham and Overton 1979, Heltshe and Forrester 1983, Smith and van Belle 1984, Chao and Lee 1990, Palmer 1990, Baltanas 1992, Mingoti and Meeden 1992, Bunge and Fitzpatrick 1993, Hodkinson and Hodkinson 1993, Colwell and Coddington 1994, Solow 1994). In particular, statistical methods derived from the capture–recapture approaches used in population biology provide useful tools for estimating species richness taking into account variation in species detectability (Burnham and Overton 1979). These methods nevertheless are seldom used in biodiversity studies.

These methods are based on a simple analogy be-

Manuscript received 25 November 1996; revised 5 May 1997; accepted 7 May 1997.

⁴ Present address: Department of Biology, Division of Zoology, P.O. Box 1050, Blindern, Oslo, 0316, Norway.

tween surveys of species occurrence and capture–recapture studies of closed animal populations used to estimate population size (Otis et al. 1978). They permit estimation of species richness when the sampling design includes at least two identified units of sampling (e.g., Burnham and Overton 1979). Records of the lists of species present during sampling sessions carried out at different times, different locations, or by different investigators may be used to estimate the species richness of a given community (Nichols and Conroy 1996). At the community level, not only will differences in the behavior of the individuals of each species introduce a potential heterogeneity in detection probability, but also the fact that the number of individuals to be potentially detected will vary among species (Nichols et al., *unpublished manuscript*). Moreover, variation in species detectability may exist among samples within a sampling session, and the detectability of the different species could be affected by changes in the behavior of the observer associated with species detection in previous samples. In the framework of the series of capture–recapture models for closed populations described in Otis et al. (1978), such different patterns of variation of species detectability may be investigated, and estimates of species richness that incorporate the patterns are available.

In this paper we demonstrate the usefulness of capture–recapture models for closed populations to estimate species richness using information from the North American Breeding Bird Survey (BBS). The BBS represents a primary source of information regarding population and community changes of land birds at a continental scale (Robbins et al. 1989, Peterjohn and Sauer 1993, Sauer et al. 1996). The BBS has a well defined sampling scheme (Robbins et al. 1986) that permits estimates of local species richness using a capture–recapture approach. In particular, each sample unit (survey route) is composed of a series of counts, each of which can be considered as a capture occasion. The BBS is used in a variety of ecological studies, and species richness is often presented as the unadjusted count of species on survey routes (e.g., Robbins et al. 1986). Because these counts are typically smaller than the number of species along routes, they are inappropriate for hypothesis tests about changes in species richness over time and space. Using selected years and regions, we documented the extent of the bias by estimating average species detection probabilities. Because of the potential importance of observer effects (Sauer et al. 1994), we evaluated whether probability of detecting species differed as a function of perceived observer ability in the survey.

We also evaluated some of the practical aspects of actually using closed-population capture–recapture estimation of species richness on the BBS data. A priori, we felt that the jackknife estimator that assumed heterogeneity in detection probabilities among species (Burnham and Overton 1979) was well suited for spe-

cies richness estimation, hence we determined how often the corresponding model was selected as most appropriate among the possible models. Finally, we investigated the potential use of the limiting form of the jackknife estimator, which uses species abundance distributions and can be used for almost any count-based sample in which numbers of individuals of each species are recorded.

These analyses permitted us to propose recommendations for the use of capture–recapture models for closed populations to estimate species richness. Although our analyses were largely exploratory in nature, we made several a priori predictions: (1) a high prevalence of heterogeneity in detection probability exists among bird species; (2) while detection probability should change with observer ability in the BBS (e.g., Sauer et al. 1994), estimated species richness values should not; (3) species richness and species detectability should vary among states; (4) jackknife estimates based on observed abundance distributions and patterns of observations among stops should be highly correlated; and, (5) that particular patterns of variation of the detection probability should occur in the data, but could be accommodated in capture–recapture models that make assumptions related to behavior, time, and individual species heterogeneity. Such expected patterns were based on hypotheses related to sampling. Specifically, in cases where “behavioral response” models were needed, we predicted that the detection probabilities for species that had not yet been observed would be lower than detection probabilities of species previously observed. This prediction is based on general ideas about observers developing a search image for particular species. In cases where models with temporal variation were selected, we looked for a pattern of decreasing detection probabilities with time elapsed since survey route initiation. This prediction is based on possible reduction in bird vocalization and activity in later morning hours. Other explanations underlying the need for models assuming a “time” effect include habitat heterogeneity producing variation in either detection of individual animals or actual species abundance.

METHODS

Survey methods and data

The BBS provided information on relative abundance of bird species at a landscape scale. The survey consisted of >4000 roadside routes located on secondary roads throughout the continental United States and Canada. Each route was 39.4 km long and surveyed once each year during the peak of breeding season. Most routes were surveyed during June. A competent observer conducted 50 3-min point counts at 0.8-km intervals on the roadside, recording all birds heard or seen during these “stops.” The survey was started in 1966 in eastern United States, and it expanded across

the continental United States and southern Canada by 1968.

Within each survey route, data can be combined differently to use various approaches to estimate local species richness. For example, BBS route data were generally summarized in five segments of 10 stops each, and we considered these as five samples from the area along the route. Species lists for each of the five groups of 10 stops for each BBS route provided the raw data used to estimate species richness with the program CAPTURE (Rexstad and Burnham 1991). The total number of individuals recorded for each species over all 50 stops was also used to build species abundance distributions. These data were used to estimate species richness using program SPECRICH (J. E. Hines et al., *in press*), which implements the limiting form of the jackknife estimator (Burnham and Overton 1979).

We investigated species richness for all routes in four states, Arizona, Maryland, North Dakota, and Wisconsin, in two years, 1970 and 1990. These states were chosen to provide contrasting habitats, bird species composition, and survey quality. Although all states have a great deal of heterogeneity in habitats, Maryland is composed primarily of various seral stages of eastern deciduous forests, Wisconsin of deciduous forest and prairie habitats, North Dakota of prairies and parklands, and Arizona of deserts and montane woodlands. The states also differed in their average number of years of survey for each route, with Wisconsin (25.5 yr/route), Maryland (23.3 yr/route), North Dakota (17.5 yr/route), and Arizona (7.7 yr/route), reflecting regional variation in availability of qualified observers. An additional source of heterogeneity in detection probability is time (Sauer et al. 1994), hence the years 1970 and 1990 were chosen to document possible changes over time in detection probabilities.

Differences among observers are a major concern in count-based surveys such as the BBS and can bias estimation of both species population change and species richness (Sauer et al. 1994). In the United States, BBS coordinators annually screen survey results for overall quality in the following way. Once routes that did not follow the prescribed methodology have been eliminated, the total numbers of species and individuals recorded are used to subjectively evaluate observer quality. Surveys by observers with unusually low counts of individuals or species are flagged to warn analysts of possible problems and to alert the observer to the need for additional training. In most published BBS results (e.g., Robbins et al. 1989), only clearly unqualified observers are omitted, as analyses of population change explicitly accommodate differences in counting ability among observers (Link and Sauer 1994). In our analyses, we used these rankings to determine whether the most competent observers had higher detection probabilities than those with lesser abilities.

Species richness estimates

Closed population models may be used to estimate species richness (i.e., the total number of species present, either within one taxon or group of species of interest, or across taxa) within a short time interval (e.g., within a given breeding season) or within a limited area in which the "population" of species is defined and not changing (Burnham and Overton 1979, Nichols and Conroy 1996). Because of this assumption of closure (the community should not change between sampling periods or locations), these models would not be applied to samples from different sampling units situated far apart (e.g., between years or between points situated far apart in space) unless it can be assumed that the probability of a local extinction or colonization occurring during the time interval, or of a strongly heterogeneous distribution of species in space at that spatial scale, is negligible. We thus assumed that each BBS route sampled a bird community to which the estimation procedure could be applied.

For the kind of community sampling schemes we considered primarily, an observer recorded presence or absence of each species of interest in each of a number of sampling units within a sampling session. Here, the sampling units were groups of stops within a survey route, but they could have been multiple counts done in the same place within a short period of time. These sampling units are analogous to the different trapping periods (e.g., each of the days on which animals are captured) of the individual capture–recapture application.

Eight closed population models seemed potentially relevant to community ecology estimation under this kind of sampling scheme, and they differed in their assumptions about sources of variation in "detection probability" (analogous to capture probability). The most simple model, denoted as M_0 , assumes that detection probability varies neither among species nor among sampling units (Otis et al. 1978:21–24). Every species has the same probability of being detected in every sampling unit. This assumption yields a multinomial model with only two parameters: total number of species, N , and detection probability, p . The maximum likelihood (ML) estimator of N under this model does not exist in a closed form and must be computed numerically. An estimator of the variance of N is also available (Otis et al. 1978). The estimator, \hat{N}_0 , is not robust to variations in detection probability and is negatively biased when detection probability varies from one species to another (heterogeneity of detection probability).

A similar but somewhat less restrictive model, M_p , permits detection probability to vary among sampling units but assumes that within each unit all species have equal detection probabilities (Darroch 1958, Otis et al. 1978:24–28). This model is characterized by $K + 1$ parameters (where K denotes the number of sampling

units; $K = 5$ for the BBS example), the total number of species, N , and the K detection probability parameters, p_j , where $j = 1, \dots, K$. The ML estimator, \hat{N}_s , must again generally be computed numerically. This estimator, \hat{N}_s , also exhibits a substantial negative bias when heterogeneity of detection probability occurs (see Otis et al. 1978).

Another somewhat less restrictive model, M_b , permits detection probability to vary by behavioral response to capture. This model may be useful in estimating species richness when the capacity for detecting a species is affected by the fact that the observer has already recorded it at a previous occasion or in a previously visited sampling unit. Parameters estimated are number of species, detection probability on any occasion of species not previously detected, and the probability of detection on any occasion for species already detected at least once (Otis et al. 1978:28–32).

The model M_{ib} assumes detection probability varies with both time and with behavioral effects. Recapture probability (denoted as c_i) is modeled as a function of initial capture probability [$c_i = p_i^{1/\theta}$], which permits all parameters to be identifiable. Parameters estimated are number of species, initial detection probabilities (p_i), and θ .

The assumption of equal detection probability among species in models M_0 , M_s , M_b , and M_{ib} will probably greatly restrict their applicability in community ecology studies. This assumption is relaxed in model M_h (Burnham and Overton 1978, 1979, Otis et al. 1978:33–37), which assumes that each species has a specific detection probability that is constant for all sampling units. Conceptually, we think of these species-specific detection probabilities, p_j (where j indexes species), as a random sample of size N from some probability distribution $F(p)$. Several frequentist approaches to estimation under model M_h have been proposed (see reviews in Bunge and Fitzpatrick 1993, Norris and Pollock 1996), and Bayesian approaches are also possible (Bunge and Fitzpatrick 1993, Solow 1994). For estimation under model M_{hs} , we seriously considered three frequentist methods: the jackknife estimator of Burnham and Overton (1978, 1979), the coverage estimators of Lee and Chao (1994), and the new nonparametric maximum likelihood estimator of Norris and Pollock (1996). Despite some concerns about its theoretical underpinnings (Bunge and Fitzpatrick 1993, Norris and Pollock 1996), the jackknife estimator tends to perform well in simulation studies (Otis et al. 1978, Burnham and Overton 1979, Pollock and Otto 1983, Norris and Pollock 1996) and in studies of known communities (Palmer 1990, 1991). Because of its performance and because of its availability in the program CAPTURE (Otis et al. 1978, Rexstad and Burnham 1991) and resultant widespread use, we selected the jackknife estimator for use with model M_h in our species richness estimation work.

The jackknife method of bias reduction (Gray and

Shucany 1972) used by Burnham and Overton (1978, 1979) leads to a series of estimators of the following form:

$$\hat{N}_l = \sum_{i=1}^K c_{il} f_i \quad (1)$$

where \hat{N}_l is the richness estimator corresponding to order l of the jackknife procedure, f_i is the number of species detected on exactly i sampling units, c_{il} are constants generated by the jackknife procedure of order l , and K is the number of sampling units (Otis et al. 1978). The estimators, corresponding estimators for the sampling variance, and a procedure for selecting the most appropriate jackknife estimator and for computing an interpolated estimator from the series of estimators of different orders are presented in Burnham and Overton (1978, 1979). The software CAPTURE implements these procedures and provides the corresponding estimates (Rexstad and Burnham 1991).

In model M_{hs} , variation in detection probability is associated with individual species and with different sampling units (Otis et al. 1978:38–40). Parameters estimated are number of species and average detection probability for each occasion (Chao et al. 1992).

The model M_{bhs} assumes detection probabilities vary by individual species and by behavioral response to detection. Parameters of this model include number of species and two probabilities of detection for each species in the community (Otis et al. 1978:40–50). No specific estimates of these detection probabilities are available for the estimation methods commonly used for model M_{bhs} . A newly developed nonparametric maximum likelihood estimator of the full M_{bhs} capture–recapture model (Norris and Pollock 1995, 1996) does permit estimation of possible sets of initial detection probability and detection probability once having been detected (i.e., estimating support points), and of their relative occurrence. This estimator is nevertheless not available in the software CAPTURE.

In the model M_{bhs} , detection probability can vary with individual species, with different sampling units and with a behavioral response to detection (Otis et al. 1978). An estimator has been recently developed for this model (Lee and Chao 1994), but it makes the additional assumption that the relative differences in detection probability among sampling units are known constants, which is not the case in this study. Moreover, this estimator is not implemented in the computer software we used (CAPTURE, Rexstad and Burnham 1991).

The investigator must decide which, if any, of the eight described models is appropriate for the data set considered. Otis et al. (1978) describe goodness-of-fit tests of specific models to the data and likelihood ratio tests between specific models and more general alternatives. They also describe a model selection procedure using a discriminant function built with simulated data.

TABLE 1. Percentage of the Breeding Bird Survey data sets for which the different species richness models were selected for each state (n = number of species). Model M_h was by far the most frequently selected model.

States	Models								n
	M_0	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbb}	
Arizona	13.9	33.3	13.9	5.6	0.0	13.9	0.0	19.4	36
Maryland	0.0	79.1	0.0	12.1	0.0	1.1	0.0	7.7	91
North Dakota	0.0	65.6	3.3	6.6	3.3	11.5	1.6	8.2	61
Wisconsin	0.0	64.8	0.8	9.4	0.8	7.0	0.0	17.2	128

Computation of estimates, model test statistics, and model selection values were all carried out using the recent update of the computer program CAPTURE (Rexstad and Burnham 1991), originally developed by Otis et al. (1978; also see White et al. 1982).

As several samples are used to estimate species richness for an area of interest, the overall detectability can be estimated as the total number of species actually recorded divided by the estimated total number of species.

The magnitude of the detection probability and the sources of variation in this parameter are extremely important in the CAPTURE models. Bias and precision of the resulting species richness estimate is directly related to this probability. Species heterogeneity, variation among sampling units, and observer behavioral effect could affect detection probability despite efforts to eliminate these sources of variation. For example, even if the investigator expends equal effort on all sampling units, it may be that species are distributed very unevenly over units as a result of microhabitat preferences, gregarious behavior patterns, or any other factor tending to produce clumped distributions of species. Moreover, activities of the birds and their detectability may decrease with the time since the survey began (surveys of BBS routes begin half an hour before sunrise; Bystrak 1981). Heterogeneity of detection probability can result from differences in population density among species at the time period or the location of interest and from interspecific differences in the probability of being heard or observed. We thus suspected that species heterogeneity was going to be the most common source of variation in detection probability and that M_h would prove to be the most useful model for estimating species richness and related parameters.

Another general sampling scheme sometimes used by community ecologists does not divide the search area into spatial sampling units or involve sampling at different times, but simply involves recording the number of individuals of each species found within a single sampling session. In this case a capture–recapture approach may also be used to estimate species richness. The data produced by this kind of sampling effort form a frequency distribution of number of individuals encountered per species. This is analogous to the frequency distribution of captures per individual in the animal population context. Burnham and Overton

(1979:934) pointed out that their model, M_h , can be made to fit this form of sampling by taking the limiting values of the coefficients c_{ij} as the number of sampling units K becomes infinite (see Eq. 1). This estimator and its relevant characteristics have already been discussed (Burnham and Overton 1979). To investigate its potential use, we computed estimates of species richness through this limiting form of the jackknife estimator using the frequency distribution of number of individuals encountered per species for a given route in a given year, and compared these estimates to those obtained using the classic jackknife estimator.

Summary of the analyses

The model selected, the significance of the goodness-of-fit test of the M_h model, estimated average species detectability under M_h , and estimated species richness under M_h were computed for each route in each state during both years. Log-linear models and ANOVAs were used to test for potential effects of state, year, and observer ability on overall detection probability and species richness. Within each state, temporal autocorrelations of average detectability and species richness were investigated at the route level. Analyses were performed using SAS (SAS Institute 1990).

RESULTS

Models selected

A total of 317 survey route-years were analyzed. The largest number of routes per state per year was in Wisconsin in 1990 ($n = 65$ route-years), the smallest in Arizona in 1970 ($n = 17$ route-years). The smallest and largest total numbers of species recorded on a route were 11 and 69 species in Arizona, 32 and 84 species in Maryland, 31 and 85 species in North Dakota, and 34 and 93 species in Wisconsin, respectively.

Globally, the M_h model was selected more often than any other of the seven alternative models (Table 1). Overall, model M_h was selected 207 times (65.3% of survey route-years analyzed). The overall proportion of goodness-of-fit tests of model M_h (against an omnibus alternative to model M_h ; Otis et al. 1978) with P values >0.05 was 63.1%. The proportion of routes for which the M_h model was selected varied among states, but not between the two years and not between the two levels of observer ability (Table 2). No pairwise interactions between year, state, and level of observer

TABLE 2. Results of log-linear model analyses testing the effects of the state, the year, and the level of observer ability on the proportion of M_h bird-species detectability models selected.

Source	df	χ^2	P
State	3	11.14	0.010
Year	1	0.56	0.455
Observer ability	1	2.85	0.092
State \times Year	3	6.34	0.096
State \times Observer ability	3	6.87	0.076
Year \times Observer ability	1	2.08	0.149
Likelihood ratio	3	2.20	0.532

Note: The tests of the effects of pairwise interactions are also reported.

ability had an effect on the proportion of routes for which the M_h model was selected (Table 2). For Arizona, M_h models were selected for a relatively small proportion of the routes (23.5% in 1970 and 40.0% in 1990), but for the three other states M_h was by far the most frequently selected model (e.g., 74.5% of the routes in 1970 and 84.1% in 1990 for Maryland).

The second most frequently selected model was M_{ibh} , and the only estimator for this model (Lee and Chao 1994) is not incorporated in program CAPTURE and requires knowledge of relative time (or sampling unit in our case) effects, which we did not have. Finally models M_{ih} and M_{bh} were selected 22 and 29 times (6.9 and 9.1% of all route-years, respectively). Overall, models including heterogeneity in the detection probabilities among individual species accounted for 94.3% of the selected models.

For routes on which model M_h was most appropriate, different hypotheses may explain the particular patterns of detection probabilities obtained (Tables 3, 4, and 5). The most likely explanation may be within-route vari-

ation in the detectability of species, which could be associated with landscape heterogeneity or with heterogeneity in the distribution of the species within the route. A trend in the detectability of the species along the five sampling segments of each route could suggest an effect of the time of day on the detectability of the birds due to a change in their activities. Nevertheless, for Arizona, North Dakota, and Wisconsin, there was a decline in the estimates of detectability with time of day on only three different routes over the 17 M_{ih} and 3 M_i selected models (see Tables 3, 4, and 5 for details about the corresponding non- M_h selected models). Conversely, the patterns of detection probabilities observed for models M_b and M_{ib} seemed to fit well with the prediction of a higher detection probability for species after its first detection (Tables 3, 4, and 5). Indeed, in seven of the eight series of estimates available for M_b and M_{ib} models selected, the estimated detection probability of the species increased after initial detection (Tables 3, 4, and 5).

Species detectabilities

Overall average bird-species detection probabilities were high but varied among states, from 0.71 for Arizona in 1990 to 0.82 for Maryland in 1970 (Fig. 1). There was no effect of year, but state and level of observer ability affected the detection probability (Table 6, Fig. 1). No pairwise interactions between these factors were significant. The overall average detection probability estimates for the most competent observers was higher (estimated detectability = 0.78 ± 0.006 [mean ± 1 SE]; $n = 268$ route-years with the most competent observers) than that for observers with lesser abilities (estimated detectability = 0.73 ± 0.016 [mean ± 1 SE], $n = 48$ route-years).

TABLE 3. Detailed estimates of the bird-species detection probabilities for non- M_h species richness models selected in Arizona.

Route number	Year	Model	Detection probability	Sampling unit				
				1	2	3	4	5
06006	1970	M_b	\hat{p}	0.68				
			\hat{c}	0.38				
06015†	1970	M_{ih}	$\hat{\bar{p}}_i$	0.36	0.29	0.30	0.18	0.16
06017	1970	M_{ih}	$\hat{\bar{p}}_i$	0.29	0.42	0.34	0.13	0.13
06021	1970	M_{ih}	$\hat{\bar{p}}_i$	0.40	0.61	0.58	0.29	0.32
06024‡	1970	M_b	\hat{p}	0.27				
			\hat{c}	0.52				
06028‡	1970	M_b	\hat{p}	0.22				
			\hat{c}	0.70				
06119‡	1990	M_b	\hat{p}	0.07				
			\hat{c}	0.45				

Notes: For model M_b , p corresponds to the initial detection probability of a species, and c to the detection probability once a species has been detected at a previous sampling unit. For model M_{ih} , the \bar{p}_i correspond to the average detection probabilities at the different sampling units. No detailed estimates of detection probability are available in software CAPTURE for previously detected species under model M_{bh} .

† Pattern of estimated detection probabilities fits the hypothesis of a decrease of detection probability with time of day.

‡ Pattern of estimated detection probabilities fits the hypothesis of a “behavioral” effect corresponding to a higher probability of detecting species after their first detection by a given observer.

TABLE 4. Detailed estimates of the bird-species detection probabilities for non- M_h models of species richness selected in North Dakota.

Route number	Year	Model	Detection probability	Sampling unit				
				1	2	3	4	5
64012	1970	M_{ih}	\hat{p}_i	0.51	0.64	0.33	0.63	0.49
64022	1970	M_{ih}	\hat{p}_i	0.33	0.35	0.46	0.22	0.31
64020	1970	M_i	\hat{p}_i	0.46	0.80	0.59	0.54	0.39
64031‡	1970	M_b	\hat{p}	0.33				
			\hat{c}	0.55				
64032	1970	M_i	\hat{p}_i	0.36	0.57	0.36	0.52	0.26
64005‡	1990	M_{ib}	\hat{p}_i	0.21	0.22	0.13	0.25	0.05
			\hat{c}_i	0.00	0.69	0.61	0.71	0.48
64007	1990	M_{ih}	\hat{p}_i	0.63	0.67	0.54	0.38	0.44
64013	1990	M_{ih}	\hat{p}_i	0.36	0.58	0.40	0.51	0.29
64026	1990	M_{ih}	\hat{p}_i	0.43	0.33	0.58	0.30	0.46
64035‡	1990	M_b	\hat{p}	0.25				
			\hat{c}	0.59				
64036	1990	M_{ih}	\hat{p}_i	0.39	0.25	0.33	0.39	0.22
64040	1990	M_{ih}	\hat{p}_i	0.44	0.37	0.62	0.42	0.28

Notes: For model M_b , p corresponds to the initial detection probability of a species, and c to the detection probability once a species has been detected at a previous sampling unit. For model M_{ib} , p_i and c_i correspond respectively to the initial detection probability of a species and to the detection probability once a species has been detected at a previous sampling unit, both at the sample unit i . For model M_{ih} , the \hat{p}_i correspond to the average detection probabilities at the different sampling units. No detailed estimates of detection probability are available in software CAPTURE for previously detected species under model M_{bh} .

‡ Pattern of estimated detection probabilities fits the hypothesis of a “behavioral” effect corresponding to a higher probability of detecting species after their first detection by a given observer.

For routes surveyed in 1970 and 1990, there was no correlation of the species detection probabilities between the two years (average detection probabilities estimated using model M_h : $r = -0.4974$, $n = 5$ route-years, $P = 0.39$ for Arizona; $r = -0.1043$, $n = 42$ route-years, $P = 0.51$ for Maryland; $r = 0.2723$, $n = 20$ route-years, $P = 0.25$ for North Dakota; $r = 0.2320$, $n = 58$ route-years, $P = 0.08$ for Wisconsin). This was also true considering only the 25 routes that were sur-

veyed both years by the same observer ($r = 0.1788$, $n = 25$ route-years, $P = 0.39$).

Species richness estimates

Species richness was different among states and between the two years considered (Table 7, Fig. 2). There was no effect of observer ability on the estimate of species richness, nor any significant interaction between state and year (Table 7). Arizona had a relatively

TABLE 5. Detailed estimates of the bird-species detection probabilities for non- M_h models of species richness selected in Wisconsin.

Route number	Year	Model	Detection probability	Sampling unit				
				1	2	3	4	5
91005‡	1970	M_b	\hat{p}	0.31				
			\hat{c}	0.60				
91011	1990	M_{ih}	\hat{p}_i	0.54	0.63	0.50	0.65	0.50
91019†	1990	M_{ih}	\hat{p}_i	0.41	0.34	0.35	0.34	0.24
91022	1990	M_{ih}	\hat{p}_i	0.43	0.61	0.65	0.48	0.36
91025	1990	M_i	\hat{p}_i	0.34	0.30	0.59	0.37	0.30
91035†	1990	M_{ih}	\hat{p}_i	0.52	0.52	0.49	0.43	0.33
91046	1990	M_{ih}	\hat{p}_i	0.50	0.30	0.54	0.32	0.33
91063	1990	M_{ih}	\hat{p}_i	0.39	0.32	0.46	0.27	0.26
91064	1990	M_{ih}	\hat{p}_i	0.45	0.67	0.69	0.46	0.63

Notes: For model M_b , p corresponds to the initial detection probability of a species, and c to the detection probability once a species has been detected at a previous sampling unit. For model M_{ih} , the \hat{p}_i correspond to the average detection probabilities at the different sampling units i . No detailed estimates of detection probability are available in software CAPTURE for previously detected species under model M_{bh} .

† Pattern of estimated detection probabilities fits the hypothesis of a decrease of detection probability with time.

‡ Pattern of estimated detection probabilities fits the hypothesis of a “behavioral” effect corresponding to a higher probability of detecting species after their first detection by a given observer.

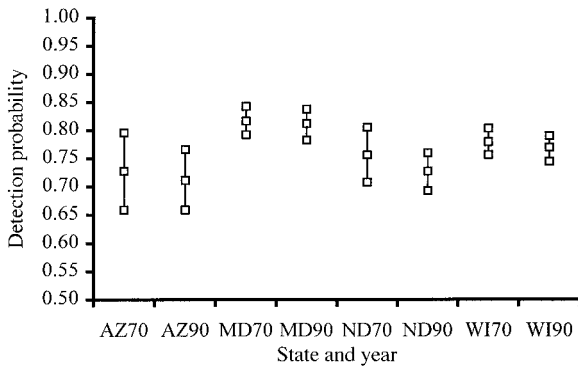


FIG. 1. Overall species detection probability (mean \pm [1.96 \times 1 SE]) for Arizona, Maryland, North Dakota, and Wisconsin, in 1970 and 1990, as estimated from Breeding Bird Survey data. AZ70 = Arizona in 1970, AZ90 = Arizona in 1990, etc.

low species richness compared to the three other states. Moreover, the average of overall species richness per route was lower in 1970 ($\hat{N} = 71.52 \pm 1.76$ species per route [mean \pm 1 SE], $n = 150$ route-years) than in 1990 ($\hat{N} = 78.43 \pm 1.64$ species per route [mean \pm 1 SE], $n = 166$ route-years).

For each state we computed the correlation of species richness on each route between the two years, and found that species richness was not correlated on each route between 1970 and 1990 for Arizona ($n = 5$ route-years; $r = 0.7655$; $P = 0.13$) or Maryland ($n = 42$ route-years; $r = 0.2164$; $P = 0.17$), but was correlated within route for North Dakota ($n = 20$ route-years; $r = 0.6240$; $P < 0.01$) and Wisconsin ($n = 58$ route-years; $r = 0.3903$; $P < 0.01$).

The jackknife species richness estimates provided by CAPTURE were highly correlated with estimates obtained by computing the limiting form of the jackknife estimator that uses observed abundance distributions of species. The distribution of the variance of the limiting form of the jackknife estimator was skewed, with most standard errors lying between 3 and 6 species. Conversely, the standard errors associated with the jackknife estimates spread largely between 3 and 14 species. There was a tendency for the jackknife esti-

TABLE 6. ANOVA results testing the effects of the state, the year, and the level of observer ability on the overall bird-species detectability p .

Source	df	Type III ss	F	P
State	3	0.2376	7.97	0.001
Year	1	0.0027	0.27	0.603
Observer ability	1	0.7550	7.59	0.006
State \times Year	3	0.0038	0.13	0.943
State \times Observer ability	3	0.0451	1.51	0.212
Year \times Observer ability	1	0.0000	0.00	0.947
Error	303	3.0121		

Note: The tests of the effects of pairwise interactions are also reported.

TABLE 7. ANOVA results testing the effects of the state, the year, and the level of observer ability on species richness estimated through model M_h (jackknife estimate).

Source	df	Type III ss	F	P
State	3	19 309.559	19.02	0.001
Year	1	2 114.327	6.25	0.013
Observer ability	1	72.582	0.21	0.644
State \times Year	3	194.398	0.19	0.902
State \times Observer ability	3	2 527.613	2.49	0.060
Year \times Observer ability	1	194.398	0.19	0.902
Error	303	146 125.402		

Note: The tests of the effects of pairwise interactions are also reported.

mate to be higher than the limiting form estimate as the mean of the difference between the two estimators computed for each route over the four states was greater than zero ($n = 316$ route-years, 4.56 ± 0.736 species per route [mean \pm 1 SE]).

DISCUSSION

In most studies, count data are used as estimates of species richness when investigating spatial or temporal trends in species richness or the effect of different environmental factors on the local occurrence of species. However to test any hypotheses about changes in species richness directly from counts of species, it must be assumed that detectability of the different species is the same for the times or locations being compared.

Using a capture-recapture approach, we showed that heterogeneity in detectability of species is highly prevalent in a survey like the BBS: heterogeneity in detectability of species was detected on most routes of the four states for the two years investigated. The overall species detection probabilities estimated using M_h models were high (range: 0.48–0.96), but varied among states. These results underline the potential biases induced by using simple counts of species as estimates of species richness, and the need to use methods that

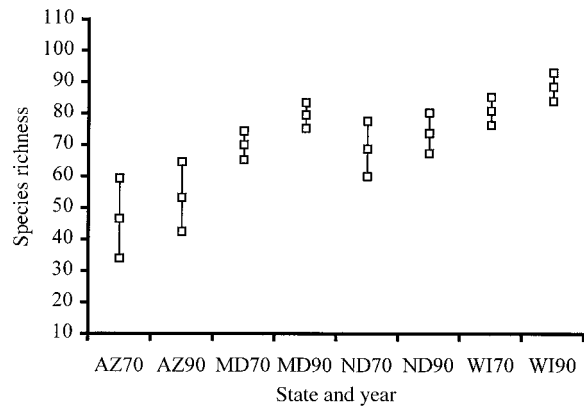


FIG. 2. Estimated bird-species richness (mean \pm [1.96 \times 1 SE]) for Arizona, Maryland, North Dakota, and Wisconsin, in 1970 and 1990, based on Breeding Bird Survey data. AR70 = Arizona in 1970, etc.

take into account heterogeneity of detection probabilities. Model M_h was by far the most frequently selected model, suggesting that variations due to time (potentially detected by M_t , M_{tb} , M_{th} , and M_{tth} models) and to a “behavioral response” (potentially detected by M_b , M_{tb} , M_{bt} , and M_{tbt} models) were not widespread. In particular, apart from Arizona, the M_h model was selected for ~70% of the routes. Detailed examination of the pattern of estimated detection probabilities of the models selected showed that the patterns of temporal and behavioral variations in detection probabilities were relatively consistent with our predictions. The selection of M_h on several routes may mean that variation in species detectability occurs among groups of stops within routes, and could be explained by heterogeneity in the landscape structure or heterogeneity in the distribution of individuals of each species in space. We did not find evidence for a “time-of-day” effect in which detection probabilities were higher for counts near sunrise than for those later in the morning. We did find evidence for a “behavioral” effect on some routes, in which the detection probability was higher for species once they had been previously detected on the route.

Our investigations showed that model M_h is the main model selected. The only state where a fair proportion of other models was selected is Arizona, but it is also the state with the lowest estimates of detection probabilities and species richness, conditions under which the model selection of CAPTURE has been shown not to perform well (Otis et al. 1978, Menkens and Anderson 1988). The assumption of heterogeneous detectability among species is thus reasonable. Moreover, simulations showed that the M_h jackknife estimator is robust to deviations from model assumptions and performs well if the average detection probability is large enough (e.g., Otis et al. 1978, Burham and Overton 1979, Pollock and Otto 1983, Norris and Pollock 1996). Finally, as using estimates obtained from different methods associated with different models (e.g., M_h : jackknife estimate; M_{bt} : removal estimate) may introduce biases in comparative tests, at this time we would recommend the use of the M_h model for large-scale analyses of data sets like the BBS. The tests using different models may nevertheless be interesting for investigating potential sources of variation in detection probabilities. It may be that proper inclusion of uncertainty associated with model selection would remove some of the problems we anticipate in using different estimators in comparative studies.

Differences in species composition and in the local abundance of species may largely explain the variation in average detectability among states. The effect of habitat structure may also directly affect the detectability of species, notably through interactions with survey methodology (B. Peterjohn, *personal communication*). For example, the BBS methodology emphasizes the detection of singing birds rather than visual

identification. In general, birds singing from the top of trees are detected over greater distances than are birds singing near the ground. Species occupying denser habitats also tend to be more vocal than those in open habitat, possibly since their risk of predation may be reduced by the dense cover. Hence, greater detectabilities for most bird species could be expected in dense landscapes rather than in open landscapes using the BBS methodology, although this would not be true for all species such as hawks that are infrequently identified by vocalizations on the BBS. In future studies, we will address this issue using analyses at the scale of the survey route to test whether landscape structure is associated with species richness and species detectability. Nevertheless, our results already suggest that factors other than landscape attributes, such as bird community composition, may strongly affect species detectability. Moreover the presence of the different species is likely to be affected by landscape structure (Flather and Sauer 1996), and thus may interact with landscape structure itself to determine the detectability of species present in a given area. Finally, observer ability can affect species detectability and, depending on the survey history in each state, this may affect the average detectability in some areas.

Although we found no evidence of differences in detection probabilities within states between 1970 and 1990, other studies have suggested that changes over time in observer ability may have a temporal component (e.g., Sauer et al. 1994). In particular, Kendall et al. (1996) have shown that start-up effects occur in BBS routes, in which lower counts tend to be associated with an observer's first year of survey along a route. We suggest that explicit considerations of detection probabilities be conducted whenever changes of species richness are evaluated over time, and computing M_h estimates for each route will take into account possible temporal variation in detection probabilities.

Species detectability varied greatly among the states we examined. It is tempting to argue by analogy that detectability of individuals within species might be reflected by the patterns among species, and that among-species detection rates could be used as ad hoc estimates of within-species detection rates. In general, analyses of BBS data are severely limited by our inability to estimate these species-specific detection rates (Barker and Sauer 1992, Flather and Sauer 1996). However, the detection rates estimated from M_h represent a mean estimate from a distribution of species-specific detection rates, hence it is unclear how the rate associated with an individual species would differ from this mean. However, if the mean detection rate is less than one (as occurred in our examples), that suggests that counts of individuals of particular species are also biased estimates of total populations and may produce misleading comparisons over time and space. Computation of trends is carried out within routes and species by species, which reduces the potential problems

associated with different detection probabilities of individuals of different species among routes. Nevertheless we have evidence that species detectability may not be correlated for a given route between two points in time situated 20 yr apart, which suggests that individual species detectability may vary in time within a route.

The pattern of species richness obtained for the four states could be expected, notably considering that we worked at the spatial scale of individual BBS routes (B. Peterjohn, *personal communication*). The relatively low species richness along BBS routes in Arizona may reflect both altitudinal stratification of bird communities in western North America, and the fact that BBS routes remain within a single physiographic stratum. Birds are more generally distributed in habitats within central and eastern North America, lacking altitudinal stratification except within a very limited number of regions, and species totals along BBS routes may tend to be more closely related to the habitat diversity along routes. The four different states were selected to exhibit different species richness in order to test the use of M_h estimators in contrasting situations. It is interesting to note a parallel increase through time in species richness for the four states considered. Particular estimators using different combinations of M_h based estimators should be used to investigate such patterns in community dynamics, and notably spatiotemporal variations in species richness or composition (e.g., Nichols et al., *in press*). In particular, potential effects of changes in landscape use and structure on community dynamics parameters could be investigated that way.

What are the implications of these findings for other types of data sets? Because many surveys contain counts collected under less standardized conditions than the BBS, it is likely that detection rates from these surveys will be even more heterogeneous than those detected in the BBS. Consequently, use of these methods is essential to ensure unbiased estimates of species richness. We suspect that model M_h will be useful in other situations when there is a sampling replication over time or space within the area under study. Numerous studies record only the number of individuals of each species in an area without replicates. In these cases, using the limiting form of the jackknife may be valuable. In the case of the BBS data considered, a good correlation between the estimates from the jackknife estimator and the limiting form of the jackknife estimator was obtained. However, sampling designs using either multiple sampling occasions or sample areas permit the testing of assumptions about sources of variation in detection probability and thus the selection of the most appropriate model.

ACKNOWLEDGMENTS

We would like to thank Bruce Peterjohn, Curt Flather, Karen McCoy, and two anonymous reviewers for their comments on previous versions of the manuscript. We greatly acknowledge the efforts of the volunteers participating in the BBS.

These highly skilled volunteers perform a great public service. This work was partly supported by the U.S. Forest Service.

LITERATURE CITED

- Baltanas, A. 1992. On the use of some methods for the estimation of species richness. *Oikos* **65**:484–492.
- Barker, R. J., and J. R. Sauer. 1992. Modelling population change from time series data. Pages 182–194 in D. R. McCullough and R. H. Barrett, editors. *Wildlife 2001: populations*. Elsevier, Amsterdam, The Netherlands.
- Bunge, J., and M. Fitzpatrick. 1993. Estimating the number of species: a review. *Journal of the American Statistical Association* **88**:364–373.
- Burnham, K. P., and W. S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**:625–633.
- Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**:927–936.
- Bystrak, D. 1981. The North American Breeding Bird Survey. *Studies in Avian Biology* **6**:34–41.
- Chao, A., S. M. Lee, and S. L. Jeng. 1992. Estimation of population size for capture–recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48**:201–216.
- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London* **B 345**:101–118.
- Darroch, J. N. 1958. The multiple-recapture census. I. Estimation of a closed population. *Biometrika* **45**:343–359.
- Dawson, D. K., J. R. Sauer, P. A. Wood, M. Berlanga, M. H. Wilson, and C. S. Robbins. 1995. Estimating bird species richness from capture and count data. *Journal of Applied Statistics* **22**:1063–1068.
- Enoksson, B., P. Angelstram, and K. Larsson. 1995. Deciduous forest and resident birds: the problem of fragmentation within a coniferous forest landscape. *Landscape Ecology* **19**:267–275.
- Flather, C. H., and J. R. Sauer. 1996. Using landscape ecology to test hypothesis about large-scale abundance patterns in migratory birds. *Ecology* **77**:28–35.
- Gray, H. L., and W. R. Shucany. 1972. *Generalized jackknife statistics*. Marcel Dekker, New York, New York, USA.
- Heltsh, J. F., and N. E. Forrester. 1983. Estimating species richness using the jackknife procedure. *Biometrics* **39**:1–11.
- Hines, J. E., T. Boulonier, J. D. Nichols, J. R. Sauer, and K. H. Pollock. *In press*. COMDYN: software to study the dynamics of animal communities using a capture–recapture approach. *Bird Study*.
- Hodkinson, I. D., and E. Hodkinson. 1993. Pondering the imponderable: a probability-based approach to estimating insect diversity from repeat faunal samples. *Ecological Entomology* **18**:91–92.
- Karr, J. R., S. K. Robinson, J. G. Blake, and R. O. Bierregaard. 1990. Birds of four neotropical forests. Pages 237–269 in A. H. Gentry, editor. *Four Neotropical rainforests*. Yale University Press, New Haven, Connecticut, USA.
- Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. *Auk* **113**:823–829.
- Knick, S. T., and J. T. Rotenberry. 1995. Landscape characteristics of fragmented shrubsteppe habitats and breeding passerine birds. *Conservation Biology* **9**:1059–1071.
- Lee, S.-M., and A. Chao. 1994. Estimating population size via sample coverage for closed capture–recapture models. *Biometrics* **50**:88–97.
- Link, W. A., and J. R. Sauer. 1994. Estimating equations estimates of trends. *Bird Populations* **2**:23–32.

- May, R. M. 1988. How many species on earth? *Science* **241**: 1441–1449.
- McIntyre, N. E. 1995. Effects of forest patch size on avian diversity. *Landscape Ecology* **10**:85–99.
- Menkens, G. E., and S. H. Anderson. 1988. Estimation of small-mammal population size. *Ecology* **69**:1952–1959.
- Mingoti, S. A., and G. Meeden. 1992. Estimating the total number of distinct species using presence and absence data. *Biometrics* **48**:863–875.
- Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. *In press*. Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conservation Biology*.
- Nichols, J. D., and M. J. Conroy. 1996. Estimation of species richness. Pages 226–234 in D. E. Wilson, F. R. Cole, J. D. Nichols, R. Rudran, and M. Foster, editors. *Measuring and monitoring biological diversity. Standard methods for mammals*. Smithsonian Institution Press, Washington, D.C., USA.
- Norris, J. L., and K. H. Pollock. 1995. A capture-recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics* **2**:305–313.
- Norris, J. L., and K. H. Pollock. 1996. Non-parametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**:639–649.
- Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**.
- Palmer, M. W. 1990. The estimation of species richness by extrapolation. *Ecology* **71**:1195–1198.
- . 1991. Estimating species richness: the second-order jackknife reconsidered. *Ecology* **72**:1512–1513.
- Peterjohn, B. G., and J. R. Sauer. 1993. North American breeding bird survey annual summary 1990–1991. *Bird Populations* **1**:1–15.
- Pollock, K. H., and M. C. Otto. 1983. Robust estimation of population size in closed animal population from capture-recapture experiments. *Biometrics* **39**:1035–1049.
- Rexstad, E., and K. P. Burnham. 1991. User's guide for interactive program CAPTURE. Abundance estimation of closed animal populations. Colorado State University, Fort Collins, Colorado, USA.
- Riffell, S. K., K. J. Gutzwiller, and S. H. Anderson. 1996. Does repeated human intrusion cause cumulative declines in avian richness and abundance? *Ecological Applications* **6**:492–505.
- Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The breeding bird survey: its first fifteen years, 1965–1979. United States Fish and Wildlife Service, Resource Publication 157, Washington, D.C., USA.
- Robbins, C. S., J. R. Sauer, R. S. Greenberg, and S. Droege. 1989. Population declines in North American birds that migrate to the neotropics. *Proceedings of the National Academy of Sciences (USA)* **86**:7658–7662.
- SAS Institute. 1990. SAS user's manual. SAS Institute, Cary, North Carolina, USA.
- Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* **111**:50–62.
- Sauer, J. R., G. W. Pendleton, and B. G. Peterjohn. 1996. Evaluating causes of population change in North American insectivorous songbirds. *Conservation Biology* **10**:465–478.
- Schiek, J., K. Lertzman, B. Nyberg, and R. Page. 1995. Effects of patch size on birds in old-growth montane forests. *Conservation Biology* **9**:1072–1084.
- Smith, E. P., and G. van Belle. 1984. Nonparametric estimation of species richness. *Biometrics* **40**:119–129.
- Solow, A. R. 1994. On the Bayesian estimation of the number of species in a community. *Ecology* **75**:2139–2142.
- Thiollay, J.-M. 1995. The role of traditional agroforests in the conservation of rain forest bird diversity in Sumatra. *Conservation Biology* **9**:335–353.
- White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture-recapture removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787-NERP.