

---

# Inference Methods for Spatial Variation in Species Richness and Community Composition When Not All Species Are Detected

JAMES D. NICHOLS,\*§ THIERRY BOULINIER,\*† JAMES E. HINES,\*  
KENNETH H. POLLOCK,‡ AND JOHN R. SAUER\*

\*U.S. Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center,  
Laurel, MD 20708, U.S.A.

†North Carolina Cooperative Fish and Wildlife Research Unit, North Carolina State University,  
Raleigh, NC 27695, U.S.A.

‡Institute of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, U.S.A.

---

**Abstract:** *Inferences about spatial variation in species richness and community composition are important both to ecological hypotheses about the structure and function of communities and to community-level conservation and management. Few sampling programs for animal communities provide censuses, and usually some species in surveyed areas are not detected. Thus, counts of species detected underestimate the number of species present. We present estimators useful for drawing inferences about comparative species richness and composition between different sampling locations when not all species are detected in sampling efforts. Based on capture-recapture models using the robust design, our methods estimate relative species richness, proportion of species in one location that are also found in another, and number of species found in one location but not in another. The methods use data on the presence or absence of each species at different sampling occasions (or locations) to estimate the number of species not detected at any occasions (or locations). This approach permits estimation of the number of species in the sampled community and in subsets of the community useful for estimating the fraction of species shared by two communities. We provide an illustration of our estimation methods by comparing bird species richness and composition in two locations sampled by routes of the North American Breeding Bird Survey. In this example analysis, the two locations (and associated bird communities) represented different levels of urbanization. Estimates of relative richness, proportion of shared species, and number of species present on one route but not the other indicated that the route with the smaller fraction of urban area had greater richness and a larger number of species that were not found on the more urban route than vice versa. We developed a software package, COMDYN, for computing estimates based on these methods. Because these estimation methods explicitly deal with sampling in which not all species are detected, we recommend their use for addressing questions about species richness and community composition.*

Métodos de Inferencia sobre Variación Espacial en la Riqueza de Especies y la Composición de Comunidades  
Cuando la Totalidad de las Especies no es Detectada

**Resumen:** *Las inferencias sobre variación espacial en la riqueza y composición de especies en comunidades son importantes tanto para el planteamiento de hipótesis ecológicas sobre la estructura y función de comunidades como para la conservación y manejo a nivel de comunidad. Pocos programas de muestreo de comunidades de animales proveen censos y usualmente algunas especies de las zonas muestreadas no son detectadas. Por lo tanto, los conteos de especies detectadas subestiman el número de especies presentes. Presentamos estimadores útiles para establecer inferencias sobre la riqueza comparativa de especies y la composición de especies entre diferentes localidades de muestreo cuando no todas las especies presentes son detectadas por los*

---

§email [jim\\_nichols@usgs.gov](mailto:jim_nichols@usgs.gov)

Paper submitted August 11, 1997; revised manuscript accepted January 22, 1998.

esfuerzos del muestreo. Nuestros métodos estiman la riqueza relativa de especies, la proporción de especies en una localidad y que son localizadas también en otra localidad y el número de especies encontradas en una localidad pero no en otra, en base a modelos de diseño robusto de captura-recaptura. Los métodos utilizan datos de la presencia y ausencia de cada especie en diferentes ocasiones de muestreo (o diferentes localidades). Para estimar el número de especies no detectadas en ninguna ocasión (o localidad). Esta aproximación permite estimar el número de especies en la comunidad muestreada, así como en fragmentos de la comunidad, mismos que son útiles para la estimación de la fracción de especies compartidas por dos comunidades. Proveemos una ilustración de nuestros métodos de estimación al comparar la riqueza de especies de aves y su composición en dos localidades muestreadas por el Estudio de Rutas de Aves Reproductoras de Norteamérica (North American Breeding Bird Survey). En este ejemplo, las dos localidades (y las comunidades de aves asociadas) representan diferentes niveles de urbanización. Estimaciones de la riqueza relativa, proporción de especies compartidas y el número de especies presentes en una ruta pero ausentes en otra, indicaron que la ruta con la fracción más pequeña de área urbana tuvo una riqueza mayor y un número mayor de especies no fueron encontradas en la ruta con más espacio urbano. Desarrollamos un paquete de cómputo, COMDYN, para determinar estimaciones basadas en estos métodos. Debido a que estos métodos de estimación pueden lidiar con muestreos en los cuales no todas las especies son detectadas, recomendamos su uso para abordar preguntas sobre la riqueza y composición de especies en comunidades.

## Introduction

Species richness is frequently viewed as a state variable for ecological systems and is often used in management and conservation efforts and in investigations of effects of habitat and human development and disturbance on biodiversity (e.g., Scott et al. 1993; Conroy & Noon 1996; Flather 1996; Keddy & Drummond 1996; Wiens et al. 1996; Worthen 1996). Comparisons of species richness estimates among different locations are used for many purposes. For example, such comparisons are used to identify potential reserves (Scott et al. 1993; Conroy & Noon 1996), to assess the relative status of different habitats (Keddy & Drummond 1996), and to draw inferences about the effects of land-use patterns (e.g., habitat fragmentation; Flather 1996) and environmental contaminants (Wiens et al. 1996) on animal communities.

In addition to comparisons of species richness among different locations, comparisons of the species compositions—(the identities of the species in a community)—of communities at different locations are also important. Estimating the proportion (or number) of species in one location that are also found in another location is relevant to many investigations in conservation biology and community ecology (e.g., when the locations differ with respect to factors such as disturbance or contamination).

Researchers often consider a species pool or source community and the proportions of species from the source that are found in various locations. Such investigations of species composition relative to a species pool are needed to test island biogeographic predictions (MacArthur & Wilson 1967; Simberloff 1974), to test ideas about community nestedness (Patterson & Atmar 1986; Lomolino 1996; Worthen 1996), and to provide metrics reflecting ecological integrity (e.g., Rapport et al. 1985).

The major methodological problem encountered when

species richness or composition is compared in different locations is the inability to count all species during sampling efforts. Early ecologists used quantitative methods to extrapolate from the number of species observed in samples to the number of species in the sampled community (e.g., Fisher et al. 1943; Preston 1948). Recently, some researchers have applied capture-recapture estimators (most originally developed to estimate animal population size) to the problem of estimating species richness (Karr et al. 1990; Bunge & Fitzpatrick 1993; Colwell & Coddington 1994; Dawson et al. 1995; Thiollay 1995; Nichols & Conroy 1996). Despite this work, however, sampling issues are frequently ignored, and it is still common in published research to see species richness—the number of species in an animal community—equated with the number of species enumerated in a sample.

Failure to detect all species in sampled communities not only leads to estimates of species richness that are negatively biased but also produces biased estimates of the comparative or relative richness of two (or more) areas. In addition, use of count statistics as naive estimates leads to problems estimating quantities reflecting differences in community composition.

We present estimation methods useful for investigating variation in species richness and composition among different locations. These estimators are based on capture-recapture models (Nichols & Conroy 1996) and use the robust design of Pollock (1982) and Pollock et al. (1990). We believe that complete censuses are extremely rare in studies of animal communities, so we developed these methods for the situation in which not all species are detected in sampling efforts. Although our focus has been on animal communities, our methods should also be useful in some studies of plant communities. Some of our estimators are spatial analogs of estimators developed to estimate quantities associated with temporal community

dynamics—variation over time rather than space (Nichols et al. 1998), and we note the appropriate analogies.

## General Sampling Design and Richness Estimation

The most common sampling design likely to be used in comparative studies of animal communities involves sampling at two different spatial scales. Such sampling can be viewed as a spatial analog of Pollock's (1982) robust design, which uses sampling at two different temporal scales. At the spatial scale at which comparisons are to be made, we sample two separate locations (Fig. 1) or perhaps one location of interest and an associated species pool (e.g., an island-mainland situation, or a single location within a region and a regional species pool). Even in the case of a specific location and a species pool, the community associated with the single location is considered a subset of but not identical to the larger community represented by the pool. In the terminology of capture-recapture models for animal populations (Pollock 1982; Seber 1982; White et al. 1982), "open" models permitting differences in the communities between the two locations are needed. We refer to the data collected at these different locations (e.g., areas A and B of Fig. 1) as "primary" samples (Pollock 1982).

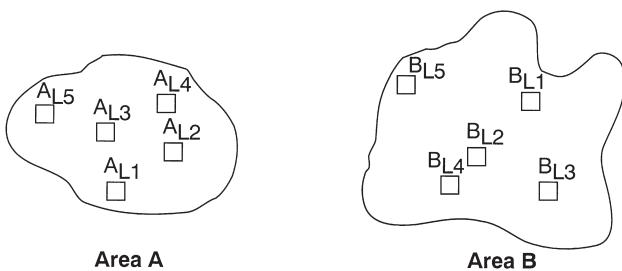
Our sampling design involves subdividing the total area within each of the two primary locations or communities into a number of secondary sampling units (quadrats) on which sampling efforts can potentially be conducted. A sample of quadrats is selected from each primary location (Fig. 1), and the investigator(s) identifies and enumerates species found on each selected quadrat using virtually any set of sampling methods, such as direct observation of animals and associated sign

(tracks, scats, nests, etc.), use of auditory cues, or trapping and netting with different trap and net types. It is best to use the same kinds of sampling methods and to expend similar effort on each of the sampled quadrats. Equal effort is not assumed or necessary, however, because we explicitly estimate area-specific species detection probabilities.

Such sampling efforts produce a list of species for each quadrat that specifies the species detected and, in some sampling situations, the number of individuals detected that belong to each species. Data collected at this scale are referred to as secondary samples, and the community is assumed to be "closed" among the different quadrats (i.e., the secondary samples are assumed to be replicate samples of the same community). Thus, the species from all quadrats within each primary sampling area will be viewed as representing the same community, and selection of primary sampling areas should reflect this view. The species list data from the secondary samples are then used to estimate species richness for the primary samples (areas A and B of Fig. 1).

Several estimators have been proposed for use with such quadrat species-list data (reviewed by Bunge & Fitzpatrick 1993). Additional estimators developed for capture-recapture data from closed animal populations (e.g., Otis et al. 1978; White et al. 1982; Rexstad & Burnham 1991) have been recommended for and used with quadrat community sampling data (Nichols & Conroy 1996; Boulinier et al. 1998; Nichols et al. 1998). Estimates based on these capture-recapture models can be computed with the software program CAPTURE (Rexstad & Burnham 1991). The richness estimates used as a basis for estimating quantities reflecting spatial variation in animal communities can be computed with any appropriate model.

In our previous work with data from the North American Breeding Bird Survey (BBS), we found strong evidence of heterogeneity in species detection probabilities (Boulinier et al. 1998). Species detection probabilities are functions of the number of individuals in a population and the behavioral characteristics of the individuals relative to sampling methods, so the finding of heterogeneity was expected. The capture-recapture model  $M_b$ , permitting heterogeneous detection probabilities among species, was usually the most appropriate model for the BBS data (Boulinier et al. 1998). For all of the example analyses we present, we used the jackknife estimator developed by Burnham and Overton (1978, 1979) for model  $M_b$ . This estimator is available in the program CAPTURE (Otis et al. 1978; Rexstad & Burnham 1991). The data required for estimation under model  $M_b$  are the total number of species detected on all quadrats and the frequencies,  $f_b$ , or numbers of species detected on exactly  $b = 1, 2, \dots, 5$  of the sampled quadrats. Other estimators for model  $M_b$  (e.g., Smith & van Belle 1984; Chao 1989; Lee & Chao 1994; Norris & Pollock 1996) and richness esti-



**Figure 1.** Generalized illustration of a sampling design permitting inferences about differences between two communities inhabiting areas A and B (primary sampling areas). Design uses spatial subsampling because quadrats ( $A_{Li}$ ,  $B_{Li}$ ) represent the secondary sampling units of the robust design. Sampling produces species lists for each quadrat, and these lists are used to compute species detection frequencies (number of species detected on exactly 1 quadrat, 2 quadrats, . . . , 5 quadrats) for each area, A and B (Table 1).

imators based on other models (e.g.,  $M_{tb}$ , incorporating both heterogeneity and time; Chao et al. 1992) can also be used with quadrat data in the community estimators we present.

## Parameters and Estimation

We use the bootstrap approach described by Nichols et al. (1998) and implemented in program COMDYN (Hines et al., in press) to estimate variances of the following estimators. Subscripts denote time and superscripts denote geographic location.

### Relative Species Richness

Relative richness, defined as the ratio of species richness for two locations, can be estimated as

$$\hat{\lambda}_i^{xy} = \frac{\hat{N}_i^y}{\hat{N}_i^x}. \quad (1)$$

The estimates of species richness for use in equation 1 ( $\hat{N}_i^y, \hat{N}_i^x$ ) are obtained by use of species occurrence data on quadrats within each location in conjunction with a species richness estimator such as the jackknife for  $M_b$ . Our purpose is not to present new estimators for species richness at specific locations but to use existing richness estimators to construct robust design estimators of quantities relevant to spatial variation in species richness and community composition.

We define  $\hat{p}_{ij}^x$  as detection probability (probability that at least one individual of species  $j$  is detected on at least one quadrat in area  $x$  during sampling efforts) for species  $j$ , given that the species is present in location  $x$  at time  $i$ . We can estimate average species detection probability for species present at location  $x$ , time  $i$ , using

$$\hat{p}_i^x = \frac{\hat{R}_i^x}{\hat{N}_i^x}, \quad (2)$$

where  $\hat{R}_i^x$  is the number of species actually detected during sampling efforts at location  $x$ , time  $i$ .

If average detection probabilities are equal for the two locations—if  $\bar{p}_i^y = \bar{p}_i^x$ —then relative species richness can be estimated more efficiently as the ratio of species observed at the two locations:

$$\hat{\lambda}_i^{xy} = \frac{\bar{R}_i^y}{\bar{R}_i^x}. \quad (3)$$

Estimators such as equation 3, based on the raw count statistics, should have smaller variances than estimators such as equation 1, which are based on estimated quantities (Skalski & Robson 1992). The estimator in equation 3, however, will be biased if detection probabilities in the two locations are not equal— $\bar{p}_i^y \neq \bar{p}_i^x$ . Under model  $M_b$ , the hypothesis of equal average species de-

tection probabilities for two samples (e.g., from locations  $x, y$ ) can be tested using the raw frequency data  $f_b$ , with a  $2 \times K$  contingency table test of the null hypothesis that the proportions of species found in  $b = 1, 2, \dots, K$  secondary samples (quadrats) are similar for the two primary sampling locations,  $x$  and  $y$ .

### Species Co-Occurrence

Many ecological questions require an assessment of similarity of the species composition of two communities. If  $\phi_i^{xy}$  is defined as the probability that a species present at location  $x$  in time  $i$  is also present at location  $y$  at that time, then one can estimate this probability as

$$\hat{\phi}_i^{xy} = \frac{\hat{M}_i^{y(R_i^x)}}{R_i^x}, \quad (4)$$

where  $\hat{M}_i^{y(R_i^x)}$  is the number of species observed in location  $x$  at time  $i$  (i.e., members of  $R_i^x$ ) that are also present in location  $y$  at that time. The statistic  $R_i^x$  is obtained directly from sampling data, so the key to equation 4 is the estimation of  $\hat{M}_i^{y(R_i^x)}$ .

The  $\hat{M}_i^{y(R_i^x)}$  can be estimated using either of two approaches, both of which focus on the species observed at location  $x$  at time  $i$  (the members of  $R_i^x$ ) and also observed at location  $y$  at time  $i$ . The most conservative approach uses only occurrence data for location  $y$  from species also detected at location  $x$  (denote these species as  $M_i^{y(R_i^x)}$ ). The occurrence data for these particular species can be used directly with one of the richness estimators to estimate  $M_i^{y(R_i^x)}$ .

The second approach requires the additional assumption that average species detection probabilities at location  $y$  are the same for the  $M_i^{y(R_i^x)}$  species that are observed in  $x$  and present in  $y$  and for the  $N_i^y - M_i^{y(R_i^x)}$  species present at location  $y$  but not observed at location  $x$ . If this assumption holds, then one can estimate the number of  $R_i^x$  also present at location  $y$  as

$$\hat{M}_i^{y(R_i^x)} = \frac{m_i^{y(R_i^x)}}{\hat{p}_i^y}, \quad (5)$$

where  $\hat{p}_i^y$  is the average detection probability of all species present at location  $y$ , time  $i$ . This average detection probability is estimated by first using occurrence data from all species observed at location  $y$  ( $R_i^y$ ) in conjunction with the program CAPTURE (Rexstad & Burnham 1991) to estimate total species richness at  $y$  ( $\hat{N}_i^y$ ) and then by using the estimator in equation 2 to estimate  $\hat{p}_i^y$ .

The first approach for estimating  $M_i^{y(R_i^x)}$  should have smaller bias but larger variance. The second approach should yield  $\hat{M}_i^{y(R_i^x)}$  with larger bias but smaller variance. A decision about the appropriateness of the second approach can be based on a test for equality of average detection probabilities for species present in primary loca-

tion  $y$  that were and were not detected in location  $x$ . This test can be conducted as a  $2 \times K$  contingency table using the  $f_b$  for the two groups of species detected in primary location  $y$ , those also detected at location  $x$  ( $m_i^{y(R_i^x)}$ ), and those not detected at location  $x$  ( $R_i^y - m_i^{y(R_i^x)}$ ). The resulting chi-squared statistic provides a test of the null hypothesis of equal proportions of species detected from the two groups. Because the group  $R_i^y - m_i^{y(R_i^x)}$  is likely to be relatively small, the contingency table test for equal detection probabilities is not likely to be very powerful, so we recommend the first approach for estimating  $M_i^{y(R_i^x)}$  for use in the estimator of equation 4.

### Number of Species at Only One of Two Locations

Some definitions of colonization rate will depend on species present at location  $y$  in time  $i$ , but not present at location  $x$  at that time ( $B_i^{xy}$ ). These species will represent potential colonists of location  $x$ . Estimation of this quantity is accomplished in one of two ways. The first approach is similar to that used to estimate the number of recruits in capture-recapture models for animal populations (Pollock 1982; Pollock et al. 1990). This approach simply subtracts from the estimated number of species at location  $y$ , the estimated number of species that occur at both locations  $x$  and  $y$ :

$$\hat{B}_i^{xy} = \hat{N}_i^y - \hat{\phi}_i^{xy} \hat{N}_i^x, \quad (6)$$

whereas the alternative estimator can be written as

$$\hat{B}_i^{xy} = (1 - \hat{\phi}_i^{yx}) \hat{N}_i^y. \quad (7)$$

We have not investigated the properties of these two estimators, so we recommend equation 6 because its capture-recapture analog (Pollock et al. 1990) is used widely and has well-known properties.

### Variation over Time and Space

Define  $\gamma_{ij}^{yx}$  as the probability that a species in species pool  $y$ , but not in location  $x$ , at time  $i$ , colonizes location  $x$  between times  $i$  and  $j$  and is still present in  $x$  at time  $j$ . This parameter can be estimated using estimates of  $B_i^{xy}$  (equations 6 & 7) and a temporal analog ( $B_{ij}^x$ ). Define  $B_{ij}^x$  as the number of species present at location  $x$  in time  $j$  ( $j > i$ ) that are "new" in the sense that they were not present at location  $x$  in time  $i$ . An estimator for  $B_{ij}^x$  is presented in Nichols et al. (1998). We can estimate  $\gamma_{ij}^{yx}$  as

$$\hat{\gamma}_{ij}^{yx} = \frac{\hat{B}_{ij}^x}{\hat{B}_i^{xy}}. \quad (8)$$

Equation 8 is based on the assumption that all new species colonizing area  $x$  between primary sampling periods  $i$  and  $j$  were members of the species pool (defined as  $y$ ) in period  $i$  (i.e., the species pool is the source of all colonizing species). Although this kind of colonization rate is relevant to a number of hypotheses in community

ecology, it may be difficult to estimate well and may tend to have a large variance.

### Example Analysis

We illustrate some of the estimators presented above using data collected as part of the BBS. This survey is carried out every spring on permanent survey routes randomly located along secondary roads throughout the United States and southern Canada. Each route is 39.4 km long and consists of 50 stops spaced at 0.8-km intervals. The observer drives along the route, exiting the vehicle at each stop to record all birds seen and heard within 0.4 km of the stop during a 3-minute observation period (Robbins et al. 1986; Peterjohn & Sauer 1993). Data in the BBS files are presently summarized by groups of 10 stops, so there are five summary records for each survey route. For our example we used the species lists for each of the five groups of stops to compute estimates on each survey route. Thus, we treated each group of 10 stops along a survey route as a "quadrat," with the five quadrats representing the secondary samples of our robust design approach. It is possible to apply our methods to any subset of total species (defined by taxonomy, foraging habit, etc.), but in our example we included all avian species.

We conducted all computations with the program COMDYN (Hines et al., in press). All of the estimators in program COMDYN are based on the standard jackknife estimators developed for model  $M_b$  by Burnham and Overton (1978, 1979). Variances are estimated in program COMDYN by a bootstrap approach (Appendix in Nichols et al. 1998). The COMDYN software includes goodness-of-fit tests of the detection frequency data to model  $M_b$  and tests of the null hypothesis that two sets of detection frequency data were produced by the same average detection probability.

We selected two BBS routes in Maryland that were fairly close geographically but that represented areas that differed substantially in degree of urbanization. We used digital land-use and land-cover data from the U. S. Geological Survey (1987) to estimate the proportion of area surrounding BBS routes that was "urban" (Flather & Sauer 1996; C. Flather personal communication). During the mid-1970s, 23% of BBS route 29 was characterized as urban, whereas only 11% of BBS route 31 was considered urban. We computed estimates of quantities reflecting geographic variation between these two routes using species detection data from 1974 (Table 1).

The species richness estimate for route 29 was about 65 avian species, whereas that for route 31 was 83 species, although the 95% confidence intervals overlapped (Table 2). The relative richness estimate,  $\hat{\lambda}_{74}^{29,31}$ , computed directly from the richness estimates was 1.28 (an estimated 28% more species on route 31 than on route

Table 1. Species detection statistics for Maryland Breeding Bird Survey routes 29 and 31 during 1974.

Species group	Species detected	Species detected on exactly h of the 5 groups of stops ( $f_h$ )				
		$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
Total species detected (rt. 29), $R_{74}^{29}$	57	13	13	8	7	16
Total species detected (rt. 31), $R_{74}^{31}$	69	19	15	17	13	5
Members of $R_{74}^{29}$ detected on rt. 31, $m_{74}^{31}(R_{74}^{29})$	50	7	9	16	13	5
Members of $R_{74}^{31}$ detected on rt 29, $m_{74}^{29}(R_{74}^{31})$	50	8	11	8	7	16

29), but the 95% confidence interval for this quantity included values slightly less than 1. The point estimates of detection probability for the two routes differed somewhat (Table 2), and the test for similar distribution of detection frequencies provided evidence of different detection probabilities ( $\chi^2_4 = 11.02, p = 0.03$ ). Because of this difference in detection probabilities, we disregarded the relative richness estimate obtained using the numbers of species observed (this also corresponds to the "naive" estimate).

The estimated fraction of route 29 species that were present on route 31 was 93%, whereas the reciprocal fraction of route 31 species found on route 29 was only 78% (Table 2). So an estimated 22% of the species present on route 31 in 1974 were not present on route 29. Similarly, the two different estimates of  $B_{74}^{29,31}$  indicated about 20 species present on route 31 that were not present on route 29, whereas fewer than 5 species were estimated to be present on route 29 and not also on route 31 (Table 2). Thus, the estimates of relative richness, proportion of shared species, and number of

species present on one route but not the other indicated that the route with the smaller fraction of urban area (route 31) had greater richness and a larger number of species that were not found on the more urban route than vice versa.

## Discussion

Spatial variation in species richness and community composition is an important factor in understanding ecological patterns and the underlying processes that produce them (Andrewartha & Birch 1954; MacArthur & Wilson 1967; Rosenzweig 1995). In many situations, however, neither species richness nor community composition is directly observable. In such situations, our ability to detect pattern and investigate process depends on reasonable estimates of the quantities of interest. We have presented estimators of community-level quantities that we believe will be useful in investigating spatial variation in species richness and community composi-

Table 2. Estimates of avian species richness, proportions of shared species, number of species in one location but not the other, and average species detection probability on two Maryland Breeding Bird Survey Routes, 29 and 31, during 1974.\*

Quantity ( $\theta$ )	Estimator	$\hat{\theta}$	$\hat{SE}(\hat{\theta})$	95% confidence interval
Species richness (rt. 29)	$\hat{N}_{74}^{29}$	65.1	5.7	57.6-81.0
Species richness (rt. 31)	$\hat{N}_{74}^{31}$	83.0	10.2	71.4-112.2
Members of $R_{74}^{29}$ present on rt. 31	$\hat{M}_{74}^{31(R_{74}^{29})}$	52.9	7.0	39.5-63.3
Members of $R_{74}^{31}$ present on rt. 29	$\hat{M}_{74}^{29(R_{74}^{31})}$	53.7	12.2	33.3-73.6
Proportion of rt. 29 species present on rt. 31	$\hat{\phi}_{74}^{29,31}$	0.93	0.10	0.69-1.00
Proportion of rt. 31 species present on rt. 29	$\hat{\phi}_{74}^{31,29}$	0.78	0.17	0.47-1.00
Relative species richness, rts. 29 and 31 (eq. 1)	$\hat{\lambda}_{74}^{29,31}$	1.28	0.19	0.99-1.74
Relative species richness, rts. 29 and 31 (eq. 3)	$\hat{\lambda}_{74}^{29,31}$	1.21	0.08	1.07-1.38
Species present on rt. 31 but not on rt. 29 (eq. 6)	$\hat{B}_{74}^{29,31}$	22.7	12.8	5.0-55.2
Species present on rt. 31 but not on rt. 29 (eq. 7)	$\hat{B}_{74}^{29,31}$	18.4	15.0	0.0-48.6
Species present on rt. 29 but not on rt. 31 (eq. 6)	$\hat{B}_{74}^{31,29}$	0.5	9.9	0.0-31.1
Species present on rt. 29 but not on rt. 31 (eq. 7)	$\hat{B}_{74}^{31,29}$	4.7	7.0	0.0-21.7
Average detection probability (rt. 29)	$\hat{p}_{74}^{29}$	0.88	0.07	0.69-0.99
Average detection probability (rt. 31)	$\hat{p}_{74}^{31}$	0.83	0.09	0.61-0.97

\*The  $\hat{\theta}$  and  $\hat{SE}(\hat{\theta})$  denote the parameter estimate and estimated standard error, respectively. Standard errors and confidence intervals are based on the bootstrap approach described in Nichols et al. (1998). All estimates were computed by COMDYN (Hines et al., in press).

tion. For example, these estimators will be useful for investigating questions related to island biogeography, metapopulation concepts, and nested subset hypotheses, as well as to conservation and management of ecosystems.

Our sampling and estimation approach is extremely flexible with respect to both sampling design and estimation models. Because of the popularity of this approach, we focused on quadrat sampling as a means of obtaining secondary samples, but secondary sampling at a particular location (i.e., within a primary sample) also can be based on sample(s) from a single study site (Bunge & Fitzpatrick 1993; Nichols & Conroy 1996). For example, in some situations a single investigator may attempt to survey an entire study area—rather than dividing it into quadrats or other sampling units—but does so on multiple occasions (Fig. 2). Data from such sampling can be expressed as species accumulation functions (Soberon & Llorente 1993; Flather 1996), and we believe that the capture-recapture model  $M_{bb}$  (Otis et al. 1978; Pollock & Otto 1983) will be useful for estimation in such situations (Nichols & Conroy 1996). In other situations, multiple investigators may visit the same area, with each investigator developing his or her own species list. In such cases, models  $M_b$  and  $M_{tb}$  (Chao et al. 1992) are likely to be most useful (Nichols & Conroy 1996). Sometimes, it may be possible at a single sampling occasion to enumerate the individuals detected in each species (Fig. 3). The resulting empirical species abundance distribution can then be used to estimate the number of species for which no individuals were observed by means of a limiting form of the jackknife estimator developed by Burnham and Overton (1979). Suggestions about the models and estimators of species

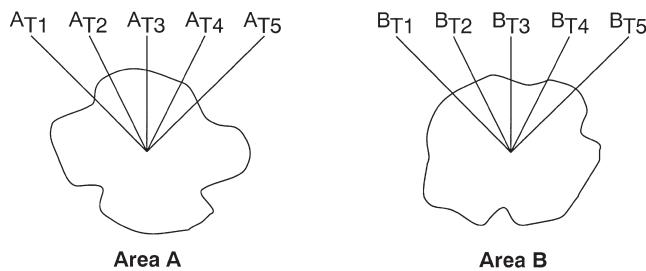


Figure 2. Generalized illustration of a sampling design permitting inferences about differences between two communities. Sampling of the communities is conducted in locations A and B. Design uses temporal subsampling; species lists are prepared on five consecutive occasions (e.g., days),  $A_{T1}, \dots, A_{T5}$  at sampling location A, and  $B_{T1}, \dots, B_{T5}$  at sampling location B. Under this design, detection frequencies are the number of species detected at a location on exactly one occasion, two occasions, . . . , five occasions.

richness likely to be most useful in these sampling situations are provided by Nichols and Conroy (1996).

For two of the quantities of interest, we showed alternative estimators. In the case of relative species richness of two locations, we adopted the approach of Skalski and Robson (1992) and recommend the estimator based on count statistics (equation 3) over the estimator based on richness estimates (equation 1) in situations for which there is no evidence of different average species detection probabilities for the two samples. Gains in precision can be substantial, and estimated standard errors and confidence intervals based on equation 1 were over twice as large as those based on equation 3 in our example. But the test for different average detection probabilities between the two samples provided evidence of different average species detection probabilities for the two areas, leading us to rely on equation 1 for the example.

Another quantity for which we provided two estimators is the number of species found in one location but not in the other. We suspected a priori that the estimator of equation 7 would tend to be more precise because it is based on two estimated quantities rather than three. In our example, however, the different estimates showed similar precision. The relative performance of these estimators is a topic that merits additional work.

The estimators we present are for parameters and other quantities that we suspect will be most useful in the study of spatial variation in animal communities. Our approach can be used to estimate other quantities as well. For example, questions involving spatial dynamics sometimes use similarity measures such as Whittaker's coefficient of community (Whittaker 1975; Farley et al.

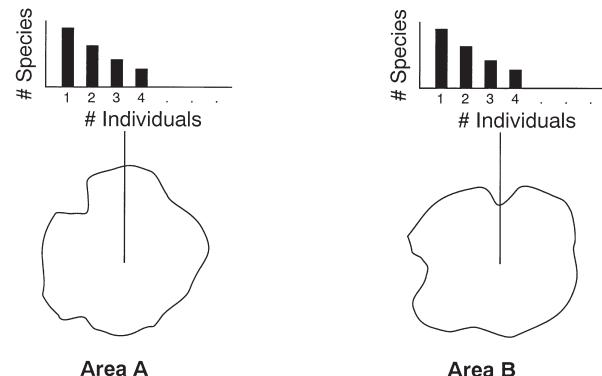


Figure 3. Generalized illustration of a sampling design permitting inferences about differences between two communities. Sampling of the communities is conducted in locations A and B but without any spatial or temporal subsampling. Instead, an empirical species abundance distribution (number of species for which exactly one individual, exactly two individuals, . . . , are detected) is obtained at each site and forms the basis for estimation using the limiting form of the Burnham and Overton (1979) jackknife estimator.

1994). This and similar metrics can be readily computed from estimates of species richness and the proportion of species present at location  $x$  that are also present at location  $y$  (equation 4). Use of capture-recapture estimates to compute such indices avoids the typical assumption of species detection probabilities of 1 and permits estimation of associated variances.

Spatial variance in species richness has been proposed as a metric useful in testing the so-called "niche limitation" hypothesis (Palmer & van der Maarel 1995; Wilson 1995) and in estimating the degree of "saturation" of communities on groups of islands (MacArthur & Wilson 1967). But measures of spatial variance based on numbers of species detected are likely to be biased because (1) the richness metrics themselves are negatively biased and (2) the variance of such metrics will reflect both true variation in richness and sampling variation associated with detection probabilities lower than 1. If richness is estimated using capture-recapture models, then the average sampling variance (Table 2) can be subtracted from the variance computed from point estimates of richness to estimate true spatial or temporal variance in richness (Skalski & Robson 1992).

Based on our previous work with BBS data sets (Boulinier et al. 1998), all estimates computed by the COM-DYN program use model  $M_b$ . Our spatial estimators, however, do not depend on any specific underlying richness estimation model. In cases where model  $M_b$  does not fit the data well, the model-fitting procedures in CAPTURE (Otis et al. 1978; Rexstad & Burnham 1991) can be used to select an alternative model and estimator. If the model selected as most appropriate by CAPTURE has available estimators and fits the data reasonably, then resulting estimates can be used with our spatial estimators. If no model fits the data well, then we recommend relying on the general robustness of the jackknife estimator for model  $M_b$  (Otis et al. 1978) and using this estimator with caution. In some cases it may be reasonable to use a quasi-likelihood approach (Burnham et al. 1987; Lebreton et al. 1992), computing variance inflation factors based on the goodness-of-fit test results.

Lack of model fit is not an adequate reason for abandoning an estimation approach and resorting to use of ad hoc estimators based on count statistics (e.g., number of species detected in a sample). Model-based estimates will likely perform much better than these ad hoc approaches even when model assumptions are not met (Nichols & Pollock 1983). Analyses that assume either that all species detection probabilities equal 1 or that species detection probabilities are the same for two different communities or habitats are not likely to lead to strong inferences (Nichols & Conroy 1996; Boulinier et al. 1998). Methods such as those proposed here deal explicitly with variable—over space, time, and species—detection probabilities and should be far preferable to ad hoc approaches for use in community investigations.

## Acknowledgments

We thank C. Flather for sharing with us the land-use and land-cover data summary used in our characterization of the two example BBS routes. We thank J. Brawn, M. Conroy, and an anonymous referee for constructive comments on the initial draft manuscript.

## Literature Cited

Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago.

Boulinier, T., J. D. Nichols, J. R. Sauer, J. E. Hines, and K. H. Pollock. 1998. Estimating species richness to make inferences in community ecology: the importance of heterogeneity in species detectability. *Ecology* 79:1018–1028.

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.

Burnham, K. P., D. R. Anderson, G. C. White, C. Brownie, and K. H. Pollock. 1987. Design and analysis methods for fish survival experiments based on release-recapture. Monograph 5. American Fisheries Society, Bethesda, Maryland.

Chao, A. 1989. Estimating population size for sparse data in capture-recapture experiments. *Biometrics* 45:427–438.

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.

Conroy, M. J., and B. R. Noon. 1996. Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. *Ecological Applications* 6:763–773.

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.

Farley, G. H., L. M. Ellis, J. N. Stuart, and N. J. Scott, Jr. 1994. Avian species richness in different-aged stands of riparian forest along the middle Rio Grande, New Mexico. *Conservation Biology* 8:1098–1108.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12:42–58.

Flather, C. 1996. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography* 23:155–168.

Flather, C. H., and J. R. Sauer. 1996. Using landscape ecology to test hypotheses about large-scale abundance patterns in migratory songbirds. *Ecology* 77:28–35.

Hines, J. E., T. Boulinier, 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*.

Karr, J. R., S. K. Robinson, J. G. Blake, and R. O. Bierregaard, Jr. 1990. Birds of four Neotropical forests. Pages 237–269 in A. H. Gentry, editor. *Four Neotropical rainforests*. Yale University Press, New Haven, Connecticut.

Keddy, P. A., and C. G. Drummond. 1996. Ecological properties for the

evaluation, management, and restoration of temperate deciduous forest ecosystems. *Ecological Applications* **6**:748-762.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**:67-118.

Lee, S. -M., and A. Chao. 1994. Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**:88-97.

Lomolino, M. V. 1996. Investigating causality of nestedness of insular communities: selective immigrations or extinctions? *Journal of Biogeography* **23**:699-703.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey.

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.

Nichols, J. D., and K. H. Pollock. 1983. Estimation methodology in contemporary small mammal capture-recapture studies. *Journal of Mammalogy* **64**:253-260.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998. Estimating rates of local extinction, colonization and turnover in animal communities. *Ecological Applications* **8**.

Norris, J. L., III, and K. H. Pollock. 1996. Nonparametric 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., and E. van der Maarel. 1995. Variance in species richness, species association, and niche limitation. *Oikos* **73**:203-213.

Patterson, B. D., and W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* **28**:65-82.

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. 1982. A capture-recapture sampling design robust to unequal catchability. *Journal of Wildlife Management* **46**:752-757.

Pollock, K. H., and M. C. Otto. 1983. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* **39**:1035-1050.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. *Wildlife Monographs* **107**.

Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* **29**:254-283.

Rapport, D. J., H. A. Regier, and T. C. Hutchinson. 1985. Ecosystem behavior under stress. *American Naturalist* **125**:617-640.

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.

Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The breeding bird survey: its first fifteen years, 1965-1979. *Resource publication 157*. U.S. Fish and Wildlife Service, Washington, D.C.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, New York.

Scott, J. M., et al. 1993. Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* **123**.

Seber, G. A. F. 1982. *Estimation of animal abundance and related parameters*. Macmillan, New York.

Simberloff, D. S. 1974. Equilibrium theory of island biogeography and ecology. *Annual Review of Ecology and Systematics* **5**:161-182.

Skalski, J. R., and D. S. Robson. 1992. *Techniques for wildlife investigations: design and analysis of capture data*. Academic Press, San Diego.

Smith, E. P., and G. van Belle. 1984. Nonparametric estimation of species richness. *Biometrics* **40**:119-129.

Soberon, M. J., and J. Llorente B. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* **7**:480-488.

Thiollay, J. M. 1995. The role of traditional agroforests in the conservation of rain forest bird diversity in Sumatra. *Conservation Biology* **9**: 335-353.

U.S. Geological Survey. 1987. Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps: data users guide. U.S. Geological Survey, National Mapping Program, Reston, Virginia.

White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture-recapture and removal methods for sampling closed populations. Publication LA-8787-NERP. Los Alamos National Laboratory, Los Alamos, New Mexico.

Whittaker, R. H. 1975. *Communities and ecosystems*. Macmillan, New York.

Wiens, J. A., T. O. Crist, R. H. Day, S. M. Murphy, and G. D. Hayward. 1996. Effects of the *Exxon Valdez* oil spill on marine bird communities in Prince William Sound, Alaska. *Ecological Applications* **6**: 828-841.

Wilson, J. B. 1995. Variance in species richness, niche limitation, and vindication of patch models. *Oikos* **73**:277-279.

Worthen, W. B. 1996. Community composition and nested-subset analyses: basic descriptors for community ecology. *Oikos* **76**:417-426.

