DISENTANGLING SAMPLING AND ECOLOGICAL EXPLANATIONS UNDERLYING SPECIES–AREA RELATIONSHIPS

EMMANUELLE CAM,[1,2,4] JAMES D. NICHOLS,[2] JAMES E. HINES,[2] JOHN R. SAUER,[2] RUSSELL ALPIZAR-JARA,[2]
AND CURTIS H. FLATHER[3]

[1]*North Carolina State University, Department of Forestry, 3024C Biltmore Hall, Raleigh, North Carolina 27695 USA*
[2]*U.S. Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center, 11510 American Holly Drive,
Laurel, Maryland 20708-4019 USA*
[3]*USDA Forest Service, Rocky Mountain Research Station, 3825 E. Mulberry Street, Fort Collins, Colorado 80524 USA*

Abstract. We used a probabilistic approach to address the influence of sampling artifacts on the form of species–area relationships (SARs). We developed a model in which the increase in observed species richness is a function of sampling effort exclusively. We assumed that effort depends on area sampled, and we generated species–area curves under that model. These curves can be realistic looking. We then generated SARs from avian data, comparing SARs based on counts with those based on richness estimates. We used an approach to estimation of species richness that accounts for species detection probability and, hence, for variation in sampling effort. The slopes of SARs based on counts are steeper than those of curves based on estimates of richness, indicating that the former partly reflect failure to account for species detection probability. SARs based on estimates reflect ecological processes exclusively, not sampling processes. This approach permits investigation of ecologically relevant hypotheses. The slope of SARs is not influenced by the slope of the relationship between habitat diversity and area. In situations in which not all of the species are detected during sampling sessions, approaches to estimation of species richness integrating species detection probability should be used to investigate the rate of increase in species richness with area.

Key words: capture–recapture; habitat diversity; nested species–area relationships; sampling artifacts; species detection probability.

INTRODUCTION

The positive relationship between the number of species and the size of the area sampled (i.e., the Species–Area Relationship, or SAR; Kinzig and Harte 2000) is a fundamental pattern in ecology that has received considerable attention (Gould 1979, Holt 1992, Rosenzweig 1995). There is an extensive literature focusing on ecological and statistical interpretations of the various models that have been used to describe these relationships (Preston 1960, 1962*a*, *b*, MacArthur and Wilson 1967, Connor and McCoy 1979, Coleman et al. 1982, Martin 1982, McGuinness 1984, Wright 1988, Lomolino 1989). In addition, SARs have played an important part in conservation biology, particularly in the SLOSS debate (i.e., a Single Large or Several Small reserves; Gilpin and Diamond 1980, Higgs and Usher 1980), and in debate about estimation of the number of species likely to become extinct as a result of land use change (e.g., deforestation) of specified magnitude (e.g., Budlansky 1994, Heywood et al. 1994, Magsalay et al. 1995, Pimm et al. 1995, Brooks and Balmford 1996, Brooks et al. 1996).

SARs are constructed by enumerating the species recorded in areas of different size. When the data are collected in smaller units of area within a larger area, the resulting relationship is called a nested SAR. Two classes of non-exclusive explanations have been proposed for the positive relationship between species richness and area in nested SARs (Preston 1960, Rosenzweig 1995, Leitner and Rosenzweig 1997). The first kind of explanation is based on sampling artifacts: some species are missed during sampling efforts, and sampling progressively larger areas and devoting more time to data collection translates into increased sampling effort (Rosenzweig 1995). This leads to a more complete list of species belonging to the community represented in the study area. The second class of explanations corresponds to ecological hypotheses, such as the influence of habitat diversity (Preston 1960, 1962*a*, *b*, Lack 1973, Freemark and Merriam 1986, Berg 1997).

The purpose of this paper is to present methods to disentangle ecological and sampling explanations underlying SARs. Specifically, our purpose is (1) to demonstrate the potential importance of sampling artifacts as determinants of SARs, and (2) to present methods to estimate SARs that reflect only ecological patterns. We have no ability to assess the relative contributions of ecological and sampling explanations to previously published SAR results, but we show how to resolve

this ambiguity in future studies. We view our methods and explanations as relevant at small and medium spatial scales, but we do not consider biogeographical explanations and processes of species origination involved at larger scales (e.g., continents; Williams 1943, Rosenzweig 1995).

Sampling artifacts and species detection probability

After reviewing much of the substantial literature on SARs, we conclude that the terms ''sampling'' and ''artifact'' have been used in different ways by different authors. For example, Connor and McCoy (1979:792–793) state that one hypothesis underlying SARs is that species number is controlled by passive sampling from the species pool, larger areas receiving effectively larger samples than smaller ones. Under this view, the sampling hypothesis is not artifactual as it is produced by a particular kind of ecological process. Indeed, the idea that the species–area relationship is purely a sampling phenomenon should be considered a null hypothesis (Connor and McCoy 1979:793). Other authors focus on the sampling by ecologists in the collection of species–area data. For example, Rosenzweig (1995:192) views as a sampling artifact the possibility that increases in diversity stem from larger sample sizes. Because of these varied views, we will try to specify the kind of sampling problem for which we propose a solution in this paper.

We view the sampling of a landscape by ecologists for the purpose of enumerating species as a probabilistic process. Define some large area of interest as well as some taxonomic group for which species richness is of interest. Assume that the ecologist goes to some smaller area within the large area of interest and uses some set of sampling methods designed to detect as many species as possible within the taxonomic group of interest. A species list is prepared of all species detected. As a way of thinking about the sampling process, focus on one particular species and assume that at least one individual of this species is present in the large area of interest. The probability of that species appearing in the ecologist's species list at the sample location can be written as the product of two conditional probabilities:

Pr(species detection|presence in large area)

= Pr(species presence in sampled area|

    presence in the large area)

  × Pr(species detection|presence in sampled area).

Some efforts to deal with sampling issues in SARs have focused exclusively on the first conditional probability (e.g., Leitner and Rosenzweig 1997). However, many discussions of sampling and SARs fail to distinguish between these two components, making it difficult to fully understand the perspectives of many authors. For example, Preston (1979) clearly recognized that not all bird species in a sampled area are detected; he recognized the second conditional probability, and that Pr[detection|presence in sampled area] <1. However, in his earlier discussions of SARs (Preston 1948, 1960, 1962*a*, *b*) in which he assumed that the overall probability of detecting a species was determined by the number of individuals in that species, it is not clear (to us) that this argument incorporated this idea of Pr(detection|presence in sampled area) <1. In this paper, when we consider detection probability and its estimation, we always include in this concept the second conditional probability, Pr(species detection|presence in sampled area). The first conditional probability is sometimes included as well, but its inclusion depends entirely on the spatial sampling design of the data collection program.

We believe that this second conditional probability is very relevant to most SARs. It is seldom possible to enumerate all of the species present in the study area, particularly when focusing on animals (Burnham and Overton 1979, Preston 1979, Nichols and Conroy 1996, Boulinier et al. 1998, 2001, Nichols et al. 1998*a*, *b*, 2000, Cam et al. 2000), but this difficulty can be encountered when studying plants as well (Connor and Simberloff 1978, Gilbert and Lee 1980, Herwitz et al. 1996). Define N to be the true number of species exposed to sampling efforts on a sample area, and S to be the number of these species that are detected in the sampling. S can be viewed as a random variable with expectation given by

$$E(S) = Np$$

where p is the conditional detection probability, reflecting the probability that a member of N appears in the count statistic, S. Detection probability is likely to be a function of many variables including effort expended in the sampling. The number of species observed in the sample, S, provides an unbiased estimate of species richness only in situations in which detection probability is equal to one. Otherwise, estimation of species richness requires use of statistical methods that explicitly incorporate species detection probability.

We are not aware of any study that has addressed detection probabilities in the framework of SARs, as most species–area curves have been fitted using count data. Lack of appropriate estimation procedures has precluded thorough investigation of the influence of detection probability on the form of SARs using empirical data. Recently developed methods permit estimation of detection probability and species richness (Burnham and Overton 1979, Boulinier et al. 1998, Nichols et al. 1998*a*, *b*). These estimates are approximately unbiased in the face of variable sampling effort, thus removing any influence of sampling effort on the estimates of richness. Sampling variances associated with these estimates will be functions of detection probability and effort (greater effort may lead to higher detection probability, which produces more precise es-

timates of richness). However, this variation in precision does not lead to biased estimates of SARs and can be dealt with via weighted analyses (using the inverse of the estimated sampling variance) if desired.

The possible confusion between the influence of sampling effort and area on species richness has long been known (e.g., Preston 1979), and several ad hoc methods have been used to attempt to disentangle these effects. The problem has sometimes been addressed in the same manner as for the influence of sample size on richness (e.g., using the rarefaction method; Sanders 1968). Sample size, effort, or sampled area may be standardized in an attempt to control for sampled area or effort (e.g., Connor and Simberloff 1978, Stevens 1986, Walter et al. 1995). A distinguishing characteristic of these approaches is that they assume some relationship between species detection probability and sampling effort, but there is frequently little empirical basis for the assumed relationship. Of course, formal methods do exist for estimating the functional relationship between sampling effort and detection probability (e.g., Pollock et al. 1984, Gould and Pollock 1987, Williams et al. 2002), and these have potential application to the estimation of species richness (Cam et al. 2001), but we have not seen these methods used to investigate SARs.

The ad hoc approaches of which we are aware do not attempt to estimate the probability that a species present in the area sampled is detected, and heterogeneity in species detection probability among species is either ignored or assumed to be directly proportional to abundance. Our intent is to use a conceptually different approach, to estimate the probability of detecting species at the level of the sampled unit of area and to explicitly incorporate this probability for estimating species richness in the framework of nested SARs. Modern statistical approaches to estimation of species richness permitted us to obtain estimates of richness that are independent of sampling effort (provided that we meet certain conditions that are specified later; see *Empirical species–area relationships*). Our general objectives are: (1) to address the contribution of increased sampling effort, and the consequent increase in detection probability, to the rate of increase in species counts with space in nested SARs; and (2) to construct SARs using species richness estimates that are unbiased in the face of variable sampling effort and that thus reflect only the ecological explanations underlying SARs.

Objectives, hypotheses, and predictions

Use of biased estimates of species richness that are dependent on effort is likely to lead to SARs whose form reflects not only the actual increase in species richness with area, but also sampling artifacts (sensu Rosenzweig 1995). Several functions can be used to describe SARs (McGuinness 1984, Flather 1996, Connor and McCoy 1979), but a model often fitted is based on the $\log_{10}$–$\log_{10}$ transformation. To avoid confusion,

we will always refer to the species–area relationship as the curve in the $\log_{10}$–$\log_{10}$ space. The corresponding model is $Y = \beta X + \alpha$, where $Y = \log_{10}$ (species richness) and $X = \log_{10}$ (number of units of area). Consequently, the rate of increase in species richness with area will correspond to the slope (β) of the linear model fitted using that transformation for both the dependent and the independent variables (i.e., richness estimates or counts, and area, respectively). Under this model, species–area relationships are usually linear (with the exception of relationships from very small areas; Rosenzweig 1995).

Nested SARs are obtained by enumerating the number of species in a unit of area (sampling unit), increasing the size of the total area sampled by sampling additional contiguous units of area, and assessing the cumulative number of species in progressively larger total areas formed by pooling a progressively larger number of contiguous units of area. Assume that the same effort is devoted to each unit of area: effort increases as the number of units sampled increases, that is, as a larger total area is sampled. As some species are missed in each sampling unit (even though they are present in the ecological community), a species is more likely to be detected as more units of area are added (that is, as larger total areas are sampled, and total effort increases). When using estimates of species richness that are not corrected for this bias (quantities such as count data that depend on total effort), the initial number of species observed in smaller areas (i.e., initial areas including a few contiguous sampling units) is likely to be smaller than the actual number of species present. Consequently, the initial number of species is likely to underestimate the actual number of species present in smaller areas, with the negative bias decreasing with increasing area sampled. Thus, the rate of increase of the cumulative number of species detected with increased area sampled is likely to be greater than the actual rate of increase in species richness with area.

The first specific objective of this paper is to present species–area curves generated by a model in which observed species richness depends exclusively on sampling effort (equal effort is devoted to each unit of area, but total effort increases as the number of contiguous units added is increased to form larger total areas). The model includes no potential influence of habitat diversity: the various species are assumed to be distributed over space in such a way that all species are available to be detected at each sampled area. This model can be viewed as a null model corresponding to the sampling component of SARs based on count data. Under this sampling model, we expect an increase in species richness with sampling effort. We present two versions of that model differing in their assumptions about detection probability (homogeneous vs. heterogeneous detection probability among species). We describe the

patterns produced by these models in a $\log_{10}$–$\log_{10}$ space.

We compare the estimates of slope parameters of SARs to those previously reported in the literature. In a large survey, Connor and McCoy (1979) provided a range of reasonable values for non-nested SARs. Kobayashi (1974) suggested that the slopes of nested SARs should be smaller than those of non-nested SARs, but provided no precise range of values. A correspondence between the characteristics of model-based curves and curves reported in previous literature on SARs would indicate that SARs based on count data from animal communities might partly reflect failure to account for species detection probability. However, whether the characteristics of model-based curves fall within a specific range corresponding to curves based on actual data is not critical here, as curves based on actual data should also reflect ecological processes not accounted for in models based on sampling effort only. We address the characteristics of model-based curves relying on various assumptions about species detection probability: a positive slope in these curves would support the hypothesis that sampling processes may contribute to shape curves based on actual data. Initially, we generate SARs based on hypothetical species detection probabilities. Then, we generate curves using estimates of detection probability corresponding to data from the North American Breeding Bird Survey (BBS; Robbins et al. 1986, Peterjohn and Sauer 1993).

Our second objective is to estimate species richness using an approach incorporating detection probability (Nichols and Conroy 1996, Boulinier et al. 1998) and to construct SARs from data on avian communities. The previously discussed modeling demonstrates the dependency of observed species richness (a biased estimate of richness) on total sampling effort (i.e., the size of the total area sampled). If detection probability is estimated and explicitly incorporated when estimating species richness, *the estimates of richness obtained are independent of the total effort devoted to an area of a given size* (i.e., a given number of contiguous units of area). As these estimates of richness corresponding to areas of different size are corrected for the proportion of species that are missed in each unit of area, they are not sensitive to the increase in the size of the total area sampled (and thus the increase in sampling effort). We compare the curves obtained using these estimates to those based on count data. Our prediction is that the slopes of curves based on the number of species observed should be steeper than those of SARs based on estimates of richness. The reasoning underlying this prediction is that variation in the cumulative number of species over space in the first case reflects both sampling artifact (the increase in sampling effort with larger area) and change in ecological conditions (e.g., change in habitat), whereas in the latter case only ecology is involved.

A final, specific objective is to use estimated species richness to test one of the ecological hypotheses put forward to explain variation in the slope of SARs: the influence of habitat diversity (Preston 1960, 1962a, Lack 1973, Freemark and Merriam 1986, Kohn and Walsh 1994, Berg 1997). The habitat diversity hypothesis assumes specialization of species on limited sets of ecological conditions. In this view, as larger areas tend to contain more habitats, they also contain more species, which would be associated with a steeper slopes for SARs. Although put forward in several papers, this hypothesis has seldom been formally addressed, and we are not aware of any study that has addressed it using approaches explicitly incorporating species detection probability for estimating species richness. We address that question using data on avian communities.

Data on Avian Communities

The BBS data are collected once a year in spring. This survey is based on a standardized sampling design: observers collect data on >4000 permanent survey routes located along secondary roads in the United States and Canada (Peterjohn and Sauer 1993). The routes are 39.4 km long, with 50 stops spaced at 0.8-km intervals. The observer drives along the route and records all the species seen or heard during a 3-min point count. This corresponds well with our view that each unit of area (the area sampled by a single point count: a circular area of radius 0.4 km) is associated with a unit of effort (3-min point count). That is, equal effort is devoted to each unit of area. We selected the data collected on BBS routes in Maryland in 1997.

Estimation of Detection Probability and Species Richness Using Avian Data

Our objectives required estimates of species richness and detection probability in order to parameterize our models and then to investigate SARs with richness estimates. We used closed capture–recapture models (Otis et al. 1978, Pollock et al. 1990) with BBS data to estimate both detection probability and species richness in bird communities (Burnham and Overton 1979, Nichols and Conroy 1996, Boulinier et al. 1998). The area of interest is divided into k sampling units ($k =$ 50 circular areas sampled by the BBS point counts) and species are enumerated in each unit. This results in a detection history for each species: a combination of k 1's and 0's reflecting the units where the species was recorded, or not, respectively.

Several closed capture–recapture models can be used to estimate species richness from detection history data (Nichols and Conroy 1996, Boulinier et al. 1998, Nichols et al. 1998a, b, Pledger 2000). The models are based on different assumptions about sources of variation in detection probability. Here we consider only two of these models: a model with homogeneous detection probability among species and sampling units, M(0),

and a model with heterogeneous detection probability among species, M(h) (Otis et al. 1978).

Under model M(0), the constant detection probability and species richness were estimated using maximum likelihood methods implemented in program CAPTURE (Otis et al. 1978, Rexstad and Burnham 1991). Under model M(h), the average detection probability and species richness were estimated using the jackknife estimators of Burnham and Overton (1978, 1979) as implemented in programs CAPTURE (Rexstad and Burnham 1991) and COMDYN (Hines et al. 1999). The jackknife estimators have performed well in simulation studies (Otis et al. 1978, Burnham and Overton 1979, Pollock and Otto 1983, Norris and Pollock 1996; R. Alpizar-Jara, J. D. Nichols, J. E. Hines, J. R. Sauer, K. P. Pollock, and C. Rosenberry, *unpublished manuscript*) and in studies of known communities (Palmer 1990, 1991). Note that we do not restrict interest to the first-order jackknife, but use the selection procedure of Burnham and Overton (1979; implemented in CAPTURE, Rexstad and Burnham 1991) to choose the proper order and compute the interpolated jackknife estimator.

In order to model heterogeneous species detection probabilities, we needed information not only about average detection probability but also about dispersion of these probabilities. When the jackknife estimator for model M(h) is used to estimate species richness when detection probability varies among species (as previously explained), only the *average* detection probability is estimated, and this approach cannot be used to assess dispersion of detection probabilities. Because we also needed information about dispersion in order to construct SAR curves under our sampling model (see *Species–area relationships generated under a sampling model*), we used a second approach to estimation of detection probabilities in the presence of heterogeneity in these probabilities among species.

We used the approach of Pledger (2000) based on a two-group distribution for capture probability. The population is assumed to be composed of two categories of individuals, each category with a specific capture probability (i.e., there is a low-p group, p^l and a high-p group, p^h). Let γ be the proportion of species in the low-p group. The mean detection probability for the community can be written as

$$\bar{p} = \gamma p^l + (1 - \gamma)p^h. \tag{1}$$

The coefficient of variation of the two-group distribution is given by

$$\text{CV}(p) = \frac{\sqrt{\gamma(1 - \gamma)(p^h - p^l)^2}}{\gamma p^l + (1 - \gamma)p^h} \tag{2}$$

(e.g., see Carothers 1973). We emphasize that use of this two-point distribution is not restrictive, because two groups are sufficient to account for heterogeneity in detection probability among species for the purpose of investigating influences on richness estimates (see Carothers 1973, Pledger 2000). The methods and associated software developed by Pledger (2000) provided estimates of the two group-specific detection probabilities (p^l and p^h), the proportion of species in the low-p group, and the CV(p) from BBS data.

To summarize, we used models M(0) and M(h) to estimate species richness from BBS data. Model M(0) was used to estimate detection probability assuming homogeneity among species. For the specific purpose of generating model-based SARs in the presence of heterogeneity in detection probability among species, we needed to estimate a dispersion parameter for these probabilities. For this purpose, we used the new estimator developed by Pledger (2000), assuming a two-group distribution for detection probabilities.

SPECIES–AREA RELATIONSHIPS GENERATED UNDER A SAMPLING MODEL

In our sampling model, we envision an area of homogeneous habitat and an associated animal community. We envision sampling effort that can be measured in discrete units, such that it is sensible to think of detection probabilities associated with each unit of effort. In community sampling, effort could be measured in a variety of ways, including quantities such as observation hours, trap-nights, or mist-net hours. However, because of our focus on area relationships, we equate one unit of area sampled with one unit of effort. Thus we assume that total sampling effort varies only with total area sampled (a given number of contiguous units of area sampled), implying equal expenditure of time and application of sampling methods on each unit of area. The homogeneity of the habitat leads to the additional assumption that a particular species is equally likely to be detected in any location and, thus, by any unit of sampling effort (i.e., the habitat is not patchy with some hot spots and with other areas where the species has no chance of being detected). Note that this assumption does not imply that each species is present at the time of sampling in each unit of area sampled, but only that a species has the same a priori probability of being present in each sampling unit regardless of the specific location of the sampling unit.

Homogeneous detection probability among species

Let p be the probability of detecting a species during a single unit of sampling effort (one unit of area), conditional on species presence in the community. Assume that this probability is the same for each species in the community. The expected proportion of species in the community detected in each unit area is p; $1 - p$ is the proportion of species missed. Let k be the number of units of area sampled. The proportion of species detected as a function of the number of units of area (equivalent to units of effort) is given by the following:

Number of units of area	Proportion of species detected
1	p
2	$p + p(1-p) = 1 - (1-p)^2$
3	$p + p(1-p) + p(1-p)(1-p) = 1 - (1-p)^3$
⋮	
k	$1 - (1-p)^k.$

If N is the total number of species present in the total area sampled, then the expected number of species detected in k units of sampling effort (denoted as S_k) is given by:

$$E(S_k) = N(1 - (1-p)^k). \qquad (3)$$

We generated species–area curves based on $k = 50$ sampling units, for consistency with the sampling design of the BBS. The total number of species (N) was arbitrarily set to 100. If p is the detection probability at the sampling unit level, then the detection probability for the total area sampled (i.e., for k sampling units) is given by

$$p_k = 1 - (1-p)^k. \qquad (4)$$

We used Eq. 3 to generate expected values of observed species richness, $E(S_k)$, for different numbers of sampling units. Expected species richness was then modeled as a linear function of the number of sampling units, k (PROC GLM, SAS; SAS Institute 1988), using the $\log_{10}$–$\log_{10}$ transformation.

Model-based curves generated using arbitrary p.— We first constructed curves corresponding to arbitrary values of detection probability selected between two extremes: 0.1 and 0.999 ($p_{50} = 0.1$ to 0.999; associated values of p range from 0.002 to 0.129). Here, p_{50} is the detection probability at the route level, which includes 50 stops (50 units of area); p is the detection probability at the stop level (i.e., the unit of area, or sampling unit). The higher the detection probability, the flatter the species–area curve, and the higher the intercept (Table 1). This reflects the positive relationship between detection probability and the proportion of species in the community collected in the first samples (i.e., in smaller total areas formed by a considering a few contiguous sampling units). Most estimates of slope parameters fall out of the range of values listed by Connor and McCoy (1979; i.e., reasonable values for non-nested SARs: 0.20–0.40). Relatively high (yet reasonable) detection probabilities (i.e., $p > 0.088$; $p_{50} > 0.99$) are needed in order to generate values in this general range (Table 1).

Model-based curves generated using estimated p.— We then selected six BBS routes in Maryland where the total number of species observed was among the highest values. We estimated detection probability under model M(0). The detection probabilities at the stop level (p) fall between 0.15 and 0.20 for the six routes

TABLE 1. Estimated parameters of species–area relationships generated under the sampling model using expected values.

p_{50}	p	$\hat{\alpha}$	$\hat{\beta}$
0.10	0.002	−0.67	0.98
0.20	0.004	−0.33	0.97
0.30	0.007	−0.12	0.95
0.40	0.010	0.05	0.93
0.50	0.014	0.19	0.90
0.60	0.018	0.33	0.87
0.70	0.024	0.47	0.83
0.80	0.032	0.62	0.79
0.90	0.045	0.81	0.71
0.91	0.047	0.83	0.70
0.92	0.049	0.85	0.69
0.93	0.052	0.89	0.68
0.94	0.055	0.92	0.66
0.95	0.058	0.95	0.65
0.96	0.062	0.98	0.63
0.97	0.068	1.03	0.61
0.98	0.075	1.09	0.58
0.99	0.088	1.18	0.53
0.995	0.100	1.25	0.49
0.999	0.129	1.38	0.41

Notes: Detection probability is homogeneous among species. The model fitted is $Y = \beta X + \alpha$, where $Y = \log_{10}$(species richness), and $X = \log_{10}$(number of units of effort). In the column heads, p_{50} is species detection probability at the route level (50 units of area); p is species detection probability at the level of one unit of area; $\hat{\alpha}$ is the estimate of the intercept parameter of the linear model; and $\hat{\beta}$ is the estimate of the slope parameter of the linear model.

(Table 2). These values lead to a detection probability approaching 1.00 at the route level (p_{50}) . Estimates of slope parameters of the model-based species–area curves generated with these values of detection probability under the sampling model (Table 2) fall within the range of reasonable values for non-nested SARs (e.g., see Connor and McCoy 1979).

Heterogeneous detection probability among species

For modeling heterogeneous detection probabilities, we followed the general approach of Carothers (1973) and considered two groups of species in each community (see *Estimation of detection probability and species richness*). If N is the total number of species

TABLE 2. Estimated parameters of species–area relationships generated under the sampling model using expected values.

Route	$\hat{p}$	$\hat{p}_{50}$	$\hat{\alpha}$	$\hat{\beta}$
150	0.1918	>0.99	1.57	0.29
51	0.1487	>0.99	1.45	0.37
34	0.1990	>0.99	1.48	0.35
17	0.1992	>0.99	1.58	0.28
11	0.1587	>0.99	1.48	0.35
07	0.1621	>0.99	1.49	0.34

Notes: Detection probability is estimated from six Maryland BBS routes using model M(0) in Program CAPTURE. The model fitted is $Y = \beta X + \alpha$, where $Y = \log_{10}$(species richness), and $X = \log_{10}$(number of units of effort). Column heads are as defined in Table 1, except that $\hat{p}_{50}$ and $\hat{p}$ are estimates.

TABLE 3.   Estimated parameters of species–area relation-
ships generated under the sampling model.

$\bar{p}_{50}$	p^{l}	p^{h}	γ	$\hat{\alpha}$	$\hat{\beta}$
A) cv(p) = 0.5					
0.49	0.012	0.013	0.9	0.13	0.91
0.60	0.005	0.029	0.3	0.44	0.82
0.72	0.012	0.044	0.4	0.62	0.76
0.81	0.001	0.318	0.2	1.66	0.17
0.99	0.075	0.226	0.5	1.42	0.37
0.995	0.096	0.257	0.8	0.92	0.47
B) cv(p) = 1.0					
0.52	0.001	0.019	0.5	0.07	0.87
0.59	0.016	0.009	0.9	0.48	0.78
0.69	0.016	0.120	0.7	0.84	0.61
0.80	0.022	0.266	0.6	1.26	0.39
0.99	0.078	0.953	0.6	1.63	0.22
0.995	0.094	0.689	0.7	1.55	0.29
C) cv(p) = 1.5					
0.52	0.010	0.160	0.8	0.76	0.58
0.59	0.016	0.177	0.9	0.61	0.70
0.73	0.024	0.264	0.9	0.78	0.65
0.79	0.001	0.202	0.7	1.05	0.31
0.99†	0.086	0.947	0.9	1.34	0.42

Notes: Detection probability is heterogeneous among spe-
cies. The model fitted is $Y = \beta X + \alpha$, where $Y = \log_{10}$(species
richness), and $X = \log_{10}$(number of units of effort). Column
heads are defined as follows: $\hat{p}_{50}$ is mean species detection
probability at the route level (50 units of area); p^{l} is detection
probability for the low-p group at the level of one unit of
area; p^{h} is detection probability for the high-p group at the
level of one unit of area; γ is the proportion of species in the
low-p group; $\hat{\alpha}$ and $\hat{\beta}$ are estimates of the intercept and slope
parameter, respectively, of the linear model.

† We could not find combinations of parameters satisfying
Eq. 2 and leading to higher values of $\bar{p}_{50}$.

present in the area of interest, then the expected number
of species detected in a sample of k units of area is
given by

$$E(S_k) = N[\gamma(1 - (1 - p^{\mathrm{l}})^k) + (1 - \gamma)(1 - (1 - p^{\mathrm{h}})^k)].$$
(5)

Model-based curves generated using arbitrary $\bar{p}$.—
We first generated three sets of curves based on three
arbitrary values of the coefficient of variation of de-
tection probability (cv(p) = 0.5, 1.0, and 1.5; Table
3). For each value of the cv(p), we computed sets of
two detection probabilities (a low p and a high p), and
the proportion of species in the low-p group (γ) to
satisfy Eq. 2. For two values of the cv(p) (0.5 and 1.0),
when the mean detection probability at the route level
($\bar{p}_{50}$) is far lower than 1, the mean detection probability
influences the shape of the species–area relationship
in the same way as for homogeneous detection probability
(i.e., the higher the detection probability, the flatter the
curve, and the higher the intercept; Table 3). However,
this association is not observed for some combinations
of γ, p^{l}, and p^{h} corresponding to very high mean de-
tection probability at the route level, or to cv(p) = 1.5
(Table 3). As with the homogeneous detection proba-
bility, most estimates of slope parameters are high rel-
ative to the values listed by Connor and McCoy (1979)

for non-nested SARs. However, curves generated on
the basis of a large difference between detection prob-
abilities of the low-p and the high-p groups lead to
estimates of slope parameters falling within the range
of reasonable values (i.e., 0.20–0.40).

Model-based curves generated using estimated $\bar{p}$.—
We then used the same six BBS routes in Maryland as
for the homogeneous detection probability examples,
and we estimated detection probability with Pledger's
(2000) model assuming two groups of species.
Pledger's approach led to a large difference between
the estimates of detection probability for the low-p and
the high-p groups at the stop level (i.e., $\hat{p}^{\mathrm{l}} < 0.10$ and
$\hat{p}^{\mathrm{h}} > 0.35$; Table 4). In addition, the estimated pro-
portion of species in the low-p group ($\hat{\gamma}$) is higher than
0.60 in all cases (Table 4). The estimated cv(p) were
slightly less than one, so our selected cv(p) in Table
3 bracketed our BBS estimates. The mean estimated
detection probability at the route level again approach-
es 1.00. As with the homogeneous detection probabil-
ity, estimates of the slope parameters of curves gen-
erated under the sampling model with heterogeneous
detection probability fall within the range of values
reported for various non-nested SARs (Connor and Mc-
Coy 1979).

Conclusion from modeling exercises

The modeling exercises reported in this section sup-
port the contention that sampling variation associated
with incomplete detection of species can produce SARs
with positive slopes. Detection probabilities estimated
from BBS data can produce SARs that have the general
appearance of those reported in the literature. We con-
clude that sampling variation associated with nonde-
tection of species and with increased sampling effort
on larger areas is probably responsible for at least some
of the shape characteristics of empirical SARs based
on count data.

EMPIRICAL SPECIES–AREA RELATIONSHIPS BASED ON COUNTS (S_k) AND ESTIMATES ($\hat{N}_k$)

To construct nested SARs using the BBS data, we
used both species counts and species richness estimates

TABLE 4.   Estimated parameters of species–area relation-
ships generated under the sampling model.

Route	$\hat{p}^{\mathrm{l}}$	$\hat{p}^{\mathrm{h}}$	$\bar{p}_{50}$	$\hat{\gamma}$	$\widehat{\mathrm{cv}}\,(p)$	$\hat{\alpha}$	$\hat{\beta}$
150	0.0677	0.3880	0.99	0.62	0.82	1.56	0.27
51	0.0782	0.4222	0.99	0.80	0.93	1.47	0.33
34	0.0909	0.4585	0.99	0.81	0.90	1.52	0.30
17	0.0652	0.4515	0.99	0.66	0.93	1.53	0.28
11	0.0594	0.4045	0.99	0.72	0.99	1.44	0.34
07	0.0623	0.3292	0.99	0.64	0.81	1.51	0.30

Notes: Detection probability is estimated from BBS data
(six Maryland routes) using Pledger's estimators (Pledger
2000). The model fitted is $Y = \beta X + \alpha$, where $Y =
\log_{10}$(species richness), and $X = \log_{10}$(number of units of ef-
fort). See Table 3 for definition of column heads; hat symbols
denote estimates, and the bar symbol ($\bar{p}_{50}$) denotes a mean
value. $\widehat{\mathrm{cv}}\,(p)$ is an estimate of the coefficient of variation of
species-level detection probability.

for groups of 4, 6, 8, 10, ..., 50 stops successively, starting with the four central stops, and successively adding one stop to each end of this central group. Sampling is conducted in a standardized manner at each stop, and equal effort is devoted to each of these units of area. This construction corresponds to the strict definition of a nested SAR: Each successively smaller area comes from within the previous larger area (Leitner and Rosenzweig 1997). In every area specified that way (even the smallest: a group of four units of area, or stops), sampling is replicated (a minimum number of four stops). Total effort increases with area (i.e., the number of replicates increases as more units of area are lumped together). However, for every area formed by lumping units of area together, we estimated species detection probability (accounting for the fact that more replicates corresponds to greater effort and, likely, higher detection probability) and corrected observed species richness by this specific detection probability. Estimates of richness obtained for areas of different size are thus independent of the total sampling effort devoted to each of these areas.

Estimation was based on model M(h), which is the model selected in the highest proportion of cases with BBS data (Boulinier et al. 1998). We used data from the 51 routes run in Maryland in 1997. We used the jackknife estimator of Burnham and Overton (1978, 1979) to estimate species richness. For each route, we modeled both the count data ($S_k = S_4, S_6, S_8, \ldots S_{50}$) and estimated species richness ($\hat{N}_k = \hat{N}_4, \hat{N}_6, \hat{N}_8, \ldots \hat{N}_{50}$) as a linear function of the number of stops, k (using the $\log_{10}$ transformation for both the dependent and the independent variables).

When modeling the species richness estimates, we could not justify the use of standard linear regression and had to think of a more general approach. Specifically, when we recognize that our $\hat{N}_k$ are estimates, rather than known values, we must consider their sampling variance–covariance structure and incorporate this into our modeling and estimation. Given that our estimates, $\hat{N}_k$, are reasonable and have small bias, then we can write the SAR estimation equation as:

$$\log_{10}\hat{N}_k = \log_{10}\alpha + \beta \log_{10}k + \varepsilon_k \qquad (6)$$

where k is the number of units of area associated with $\hat{N}_k$, and ε_k is a random error term with expectation $E(\varepsilon) = 0$ (the estimators, $\hat{N}_k$, should have small bias), and covariance matrix $E(\varepsilon\varepsilon') = \Sigma$. Because the route segments are nested, with data from smaller segments also used to estimate richness on larger segments, we expect a covariance between any pair of richness estimates, $\hat{N}_i$ and $\hat{N}_j$. In order to estimate this sampling covariance, we follow the approach of Otis et al. (1978:69–70) in their analysis of nested subgrids for density estimation from capture–recapture data. We assume that the sampling correlation between $\hat{N}_i$ and $\hat{N}_j$ is equal to the proportion of overlapping area between the two route seg-

ments (computed as i/j where $i < j$) and then estimate $\text{cov}(\hat{N}_i, \hat{N}_j)$ as

$$\text{cov}(\hat{N}_i, \hat{N}_j) = \widehat{\text{SE}}(\hat{N}_i) \times \widehat{\text{SE}}(\hat{N}_j) \times \text{corr}(\hat{N}_i, \hat{N}_j)$$

$$= \widehat{\text{SE}}(\hat{N}_i) \times \widehat{\text{SE}}(\hat{N}_j) \times (i/j) \qquad (7)$$

for $i < j$. The $\widehat{\text{SE}}(\hat{N}_k)$ come directly from our jackknife estimation algorithm (Burnham and Overton 1978, 1979). Generalized linear least squares is then used with the $\log_{10}$-transformed data in order to directly estimate model parameters α and β under the model of Eq. 6.

We estimated the slope parameters of species–area relationships based on counts on the one hand, and on estimates on the other hand. In 45 routes out of 51, the slope of the relationship based on counts was steeper than that of the curve based on estimates (Table 5). A common relationship was fit to the data from all routes pooled. The estimated slope parameter of the relationship based on counts was higher than that of the relationship based on estimates of species richness (Fig. 1). We also used the difference between estimates and counts (which estimates the species not detected) and addressed the relationship between that quantity and the number of sampling units (model built: $Y = \beta X + \alpha$, where $Y = \log_{10}(\hat{N}_k - S_k)$ and $X = \log_{10}(k)$). Our prediction was that this difference should decrease as the number of sampling units increases, and the results were consistent with this prediction ($F_{1, 1222} = 329.26$, $P < 0.0001$, $\hat{\alpha} = 0.267$, $\widehat{\text{SE}} = 0.009$; $\hat{\beta} = -0.116$, $\widehat{\text{SE}} = 0.006$). The results of our empirical analyses followed our prediction that SARs based on raw species counts would reflect both sampling artifact and true ecological explanations and would thus be steeper than SARs that were based on estimates (counts corrected for effort) that would thus reflect only ecological explanations.

INFLUENCE OF HABITAT HETEROGENEITY ON THE SLOPE OF SARs

To address the question of the influence of habitat heterogeneity on the slope of SARs, we used GIS to specify a circular buffer of radius 0.40 km centered on each stop of nine Maryland BBS routes run in 1997 for which individual stops have been digitized. We used the coverage developed by Jones et al. (1997) from data collected by satellite to assess habitat diversity. In this coverage, habitat is classified into 15 categories (e.g., mixed, deciduous, or conifer forest, row crop, woody wetlands, high- and low-intensity developed, etc.). We characterized habitat diversity in buffers centered on groups of stops. As for species richness, we started with the four central stops, then progressively added two circular buffers at each end of this group, and assessed habitat diversity in the total area corresponding to the juxtaposition of buffers.

We assessed habitat diversity in two ways. We first used Shannon's Diversity Index (McGarigal and Marks

TABLE 5. Estimates of slope parameters of species–area curves constructed from count data ($\hat{\beta}_c$) and from estimates ($\hat{\beta}_e$) of species richness (Model M(h)) corresponding to 51 BBS routes in Maryland.

Route	Count data		Estimates		
	$\hat{\beta}_c$	$\widehat{SE}(\hat{\beta}_c)$	$\hat{\beta}_e$	$\widehat{SE}(\hat{\beta}_e)$	$\hat{\beta}_c - \hat{\beta}_e$
1	0.55	0.03	0.39	0.12	0.16
3	0.40	0.01	0.23	0.08	0.17
4	0.57	0.02	0.55	0.08	0.02
5	0.27	0.01	0.16	0.04	0.11
6	0.24	0.01	0.09	0.06	0.15
7	0.36	0.01	0.27	0.09	0.09
8	0.28	0.01	0.12	0.05	0.16
9	0.22	0.01	0.09	0.05	0.13
11	0.45	0.02	0.47	0.12	−0.02
12	0.43	0.02	0.45	0.07	−0.02
13	0.50	0.02	0.46	0.17	0.04
14	0.33	0.01	0.28	0.08	0.05
15	0.45	0.01	0.62	0.14	−0.17
16	0.21	0.01	0.02	0.09	0.19
17	0.31	0.01	0.21	0.09	0.10
18	0.55	0.03	0.66	0.10	−0.11
19	0.37	0.02	0.37	0.12	0.00
20	0.45	0.01	0.43	0.08	0.02
22	0.46	0.02	0.49	0.10	−0.03
24	0.30	0.03	0.11	0.08	0.19
27	0.42	0.01	0.34	0.06	0.08
28	0.30	0.02	−0.05	0.12	0.35
29	0.35	0.01	0.20	0.11	0.15
30	0.46	0.01	0.26	0.12	0.20
31	0.25	0.02	0.02	0.06	0.23
32	0.40	0.01	0.33	0.14	0.07
33	0.35	0.01	0.17	0.07	0.18
34	0.35	0.02	0.14	0.06	0.21
36	0.37	0.01	0.20	0.08	0.17
37	0.29	0.02	0.25	0.08	0.04
38	0.43	0.02	0.31	0.07	0.12
39	0.33	0.01	0.26	0.12	0.07
40	0.28	0.01	0.32	0.08	−0.04
44	0.38	0.01	0.29	0.10	0.09
45	0.36	0.02	0.25	0.09	0.11
46	0.44	0.01	0.34	0.11	0.10
47	0.32	0.01	0.20	0.09	0.12
48	0.34	0.02	0.09	0.09	0.25
51	0.42	0.02	0.27	0.10	0.15
52	0.60	0.02	0.39	0.15	0.21
53	0.48	0.02	0.28	0.11	0.20
54	0.31	0.02	0.16	0.06	0.15
55	0.46	0.03	0.17	0.10	0.29
102	0.44	0.01	0.27	0.09	0.17
110	0.33	0.01	0.18	0.01	0.15
112	0.29	0.01	0.28	0.11	0.01
121	0.39	0.01	0.29	0.09	0.10
123	0.21	0.01	−0.01	0.10	0.22
125	0.30	0.01	0.18	0.06	0.12
126	0.28	0.03	0.07	0.07	0.21
150	0.30	0.01	0.04	0.07	0.26

Notes: The model fitted is $Y = \beta X + \alpha$, where $Y = \log_{10}$ (species richness), and $X = \log_{10}$ (number of stops). $\widehat{SE}$ is the estimated standard error.

1994), which is based on the proportion of patches of each category. Here a patch corresponds to a pixel in the digital map. That index increases when habitat diversity increases. It has no upper limit, but approaches zero when there is no diversity. We also used the number of land types in the area of interest (i.e., the buffers centered on the stops in a group of k stops) as an alternative diversity metric.

We estimated the slope of relationships between one of the measures of habitat diversity (i.e., Shannon's Diversity Index or the number of land types) and area (i.e., the number of stops taken into account). To test the hypothesis of an influence of habitat diversity on the rate of increase of species richness with area, we modeled the slope of species–area relationships as a function of the slope of habitat–area relationships (PROC GLM, SAS 1988). By using this approach, it is the actual rate of increase in diversity obtained by adding units of area that may include different habitat types (not previously encountered) that is assessed. An alternative approach would rely on assessment of the overall habitat diversity at the scale of the larger area considered, and comparison among slopes of SARs (for bird communities) corresponding to different overall levels of habitat diversity. However, the same overall level of diversity may result from very different spatial arrangements of the patches of different types: diversity may be high only at one end of the total area considered. We believe that the slope of the habitat–area relationship is more relevant for the purpose of testing the hypothesis of a relationship between changes in habitat diversity and the slope of species–area relationship in bird communities. Using this approach to quantify habitat diversity, if bird species actually specialize on a specific set of habitat types, we might expect a positive relationship between the rate of increase in habitat diversity on the one hand, and in the number of bird species on the other hand, with area.

We wanted to restrict inference to ecological hypotheses, so we used estimates of species richness (not observed richness) computed using the jackknife estimator of Burnham and Overton (1979; see *Estimation of detection probability and species richness using avian data*). We estimated the slope of species–area relationships for the nine routes whose stops have been digitized.

Under the hypothesis of a higher rate of increase in species richness with area in cases in which habitat diversity increased more quickly with area, we might expect a positive relationship between the slope of the species–area relationship and that of the habitat–area relationship. We did not find evidence of an influence of the slope of the habitat–area relationship based on Shannon's Index on the slope of species–area relationships ($F_{1,7} = 0.13$, $P = 0.73$, $R^2 = 0.02$). Results were similar using the slope of the habitat–area relationship based on the number of land types ($F_{1,7} = 0.25$, $P = 0.63$, $R^2 = 0.03$).

DISCUSSION

Species–area relationships have long been present in the literature on animal and plant communities (Preston 1960, Kilburn 1966, MacArthur and Wilson 1967), and they still form the basis of current theoretical and empirical studies (Caswell and Cohen 1993, Haydon et al. 1993). Here we consider a specific kind of sampling

Fig 1. Species–area relationships based on count data and estimates of species richness [model M(h)] from 51 BBS routes in Maryland, USA. Expected values are based on a linear model. The model fitted is $Y = \beta X + \alpha$, where $Y = \log_{10}(\text{species richness})$ and $X = \log_{10}(\text{number of units of area})$.
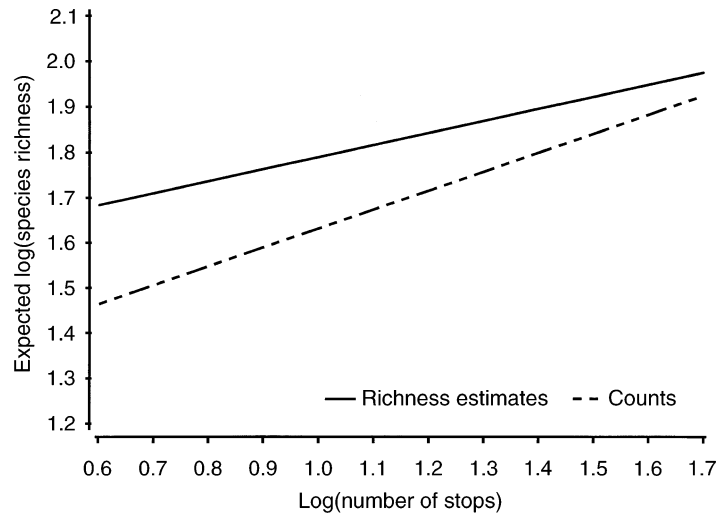
artifact (some species present in sampled areas are missed during sampling efforts) and investigate the influence of this artifact on species–area analyses.

Our results indicate that accounting for species detection probability influences the slope of species–area relationships in a predictable manner: the slope of SARs based on counts is steeper than that of SARs based on estimates of species richness (in a $\log_{10}$–$\log_{10}$ space). This is exactly what would be predicted if both sampling and ecological processes contribute to SARs based on count data. In addition, species–area curves generated under sampling models can look realistic: their slope parameters can fall within the range of estimates reported in the large survey of Connor and McCoy (1979). These authors, however, addressed non-nested SARs. Kobayashi (1974) suggested that the slopes of nested SARs (continuous sampling; Kobayashi 1974, 1975) should be smaller than those of non-nested SARs (discrete sampling; Kobayashi 1974). However, we are not aware of any published work reporting ranges of values expected for nested SARs. Many of the values that we obtained using arbitrary detection probabilities (assuming either homogeneity or heterogeneity in detection probability) led to very large slope parameters for model-based SARs. As expected, the smaller the detection probability, the larger the slope parameter. However, the slopes of model-based SARs generated using values estimated from BBS data were closer to values reported in the literature.

Under our sampling models, the increase in species counts with space depends on sampling effort exclusively: these curves do not reflect any potential ecological component underlying SARs. The sampling process underlying these model-based curves is likely to contribute to the form of relationships based on species counts. This process will not be reflected in relationships based on estimates of richness, because estimates can be viewed as counts "corrected for effort" and are independent of total effort devoted to sampling (and thus of the size of the area sampled).

Tests of hypotheses concerning ecological processes underlying SARs are possible on condition that one has access to the increase in richness with space that is exclusively explained by these processes. Estimates of richness obtained using capture–recapture models lead to species–area curves that do not include the sampling component of variation present in curves based on count data. This permitted us to test one of the hypotheses proposed to explain variation in the slope of SARs: the influence of habitat diversity. We addressed that question by investigating the relationship between the rate of increase in habitat diversity with area and the rate of increase in species richness with area. Our results do not provide evidence that the slope of the habitat–area relationship influences the slope of SARs. Of course it is possible that these results reflect use of categories of habitat or of an index of habitat diversity that do not contain the features structuring bird communities and determining local species richness. Even if the selected habitat components are indeed relevant to the structuring of bird communities, it is possible that the mobility of birds does not permit the discrimination of structure at this geographic scale with this type of sampling. Lastly, higher habitat diversity implies less area per habitat: the hypothesis of a positive relationship between the slope of SARs and the degree of habitat diversity should probably be assessed not only in the light of the hypothesis of habitat specialization, but also of area requirements. Further work is needed to investigate that question more thoroughly. If data are available, the approach presented here can be used to explore the relationship between the rate of increase in species richness in communities defined ecologically (e.g., groups of species such as forest species) with space and the rate of increase in habitat

diversity with space using highly stratified data on habitat (e.g., different types of forests, characteristics of trees, etc.).

Early work on the contribution of ecology and sampling processes to the form of SARs noted the potential importance of sampling artifacts (Williams 1943, Preston 1960). This motivated several attempts to address the increase in richness not linked to collection of larger samples (i.e., to increased sampling effort; reviewed by Rosenzweig 1995). Rarefaction (Sanders 1968, Simberloff 1972, Heck et al. 1975) has been recommended as a means of comparing species richness of two sampled communities. The approach is based on the assumption of similar species abundance distributions for the communities to be compared. The proportion of total community individuals in each species is estimated using the larger of the two community samples, and the expected number of species detected in the smaller sample of n individuals is then computed from these estimates. Sanders (1968) recommended rarefaction in comparisons involving similar sampling methods, similar taxa, and similar habitats (also see Simberloff 1979). These restrictions seem to be related to the assumptions of similar species abundance distributions and similar individual-level detection probabilities for members of any species in the two different locations or communities. Our approach of using probabilistic estimators for species detection probability and species richness does not require these restrictive assumptions and does not require a close correspondence between the number of individuals in a species and its detection probability (although such a relationship is likely to exist; R. Alpizar-Jara, J. D. Nichols, J. E. Hines, J. R. Sauer, K. P. Pollock, and C. Rosenberg, *unpublished manuscript*). Thus far, debates concerning the mathematics of species–area curves have focused on the form of the relationship, hence, on the equation that should be used (Connor and McCoy 1979), and on the underlying species abundance distribution (Connor and McCoy 1979, Coleman 1981, McGuinness 1984). They have not focused on the influence of the sampling process itself on the shape of curves.

Our results indicate that the rate of increase in species richness over space (using the $\log_{10}$–$\log_{10}$ transformation) can be overestimated when using count data. The contribution of sampling artifacts to the increase in species richness based on empirical data is likely to vary according to the kind of organisms studied and the design of the study. Lists of species from long-term survey programs with repeated collection of data in each location, for example, may approximate true species richness well and thus mitigate the contribution of increased sampling effort with area to the increase in richness over space. However, it is not always possible to take advantage of cumulative field effort over time to obtain a more accurate estimate of richness. The objectives in conservation biology might

sometimes be incompatible with the use of data collected repeatedly over long periods of time. Assessment of the contemporary state of ecological systems may require data collected within short periods of time. In addition, data from long-term survey programs can be used to explore variation of species richness with space on condition that one assumes that local extinction and colonization events in the area of interest are negligible during the period of time over which the data are collected. If long-term data are not used, and in situations in which not all of the species are detected during sampling efforts, count data are likely to lead to underestimation of species richness. Approaches to estimation of species richness should be used for further investigations of factors influencing the characteristics of SARs. Estimates can certainly also be used to address the form of the relationship between species richness and area (e.g., Connor and McCoy 1979). This is beyond the scope of this paper.

Although the modeling and estimation presented here have concerned nested SARs, we believe that probabilistic estimates of species richness should be used to investigate non-nested SARs as well. There is large potential for differing detection probabilities in studies of non-nested SARs for a variety of reasons, including different observers and different habitats as well as differing amounts of sampling effort. Because estimates of species richness permit different detection probabilities for different locations, this approach should also be used to address non-nested SARs. Obtaining reliable estimates of richness that do not reflect sampling processes is a step that should precede investigations of the relevance of the various mathematical functions that could be used to describe the shape of SARs. Hypotheses about ecological processes assumed to influence the shape of SARs (e.g., reviewed in Connor and McCoy 1979) should also be tested using approaches that disentangle true variation in species richness over space and variation resulting from sampling processes.

The sampling model developed here describes the increase in species richness linked to the increase in sampling effort exclusively: this does not involve any ecological process. In their important paper on the statistics and biology of the SAR, Connor and McCoy (1979) stated that the idea that the species–area relationship is purely a sampling phenomenon should be considered as a null hypothesis; all hypotheses invoking biological processes to explain the species–area relationship would be considered alternatives. McGuinness (1984) noted that "They did not, however, amplify this, nor did they describe analytical tools that would make a statistical test of this hypothesis possible." The sampling model presented here differs from that described by Connor and McCoy (1979): the phenomenon that they invoked corresponds to passive sampling from the species pool (larger areas are supposed to receive larger samples and, consequently, more spe-

cies). However, they have in common the underlying idea of assessing the rate of increase in species richness with space under the assumption that no (or minimal) ecological process is involved. Under this view, the sampling model presented here can be considered as a null model for use in studies based on count data, which is the case in the vast majority of previous studies. However, for future work we recommend use of estimation methods of species richness integrating species detection probability.

Literature cited

Berg, Å. 1997. Diversity and abundance of birds in relation to forest fragmentation, habitat quality and heterogeneity. Bird Study **44**:355–366.

Boulinier, T., J. D. Nichols, J. E. Hines, J. R. Sauer, C. H. Flather, and K. H. Pollock. 2001. Forest fragmentation and bird community dynamics: inference at regional scales. Ecology **82**:1159–1169.

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

Brooks, T., and A. Balmord. 1996. Atlantic forest extinctions. Nature **380**:115.

Brooks, T. M., S. L. Pimm, and N. J. Collar. 1996. Deforestation predicts the number of threatened birds in insular Southeast Asia. Conservation Biology **11**:382–394.

Budlansky, S. 1994. Extinction or miscalculation. Nature **370**: 105.

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.

Cam, E., J. D. Nichols, J. R. Sauer, and J. E. Hines. 2001. On the estimation of species richness based on the accumulation of previously unrecorded species. Ecography (*in press*).

Cam, E., J. D. Nichols, J. R. Sauer, J. E. Hines, and C. H. Flather. 2000. Relative species richness and community completeness: avian communities and urbanization in the mid-Atlantic states. Ecological Applications **10**:1196–1210.

Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. Biometrics **29**:79–100.

Caswell, H., and J. E. Cohen. 1993. Local and regional regulation of species–area relations: a patch-occupancy model. Pages 99–107 *in* R. E. Ricklefs and D. Schluter, editors. Species diversity in ecological communities. University of Chicago Press, Chicago, Illinois, USA.

Coleman, B. D. 1981. On random placement and species–area relations. Mathematical Biosciences **54**:191–215.

Coleman, B. D., M. A. Mares, M. R. Willig, and Y.-H. Hsieh. 1982. Randomness, area, and species richness. Ecology **63**: 1121–1133.

Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species–area relationship. American Naturalist **113**:791–833.

Connor, E. F., and D. S. Simberloff. 1978. Species number and compositional similarity of the Galapagos flora and fauna. Ecological Monographs **48**:219–248.

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

Freemark, K. E., and H. G. Merriam. 1986. Importance of area and habitat heterogeneity to bird assemblages in temperate forest fragments. Biological Conservation **36**:115–141.

Gilbert, F. S., and S. B. Lee. 1980. Two perils of plant population dynamics. Oecologia **46**:283–284.

Gilpin, M. E., and J. M. Diamond. 1980. Subdivision of nature reserves and the maintenance of species diversity. Nature **285**:567–568.

Gould, S. J. 1979. An allometric interpretation of species–area curves: the meaning of the coefficient. American Naturalist **114**:335–343.

Gould, W. R., and K. H. Pollock. 1997. Catch-effort maximum likelihood estimation of important population parameters. Canadian Journal of Fisheries and Aquatic Sciences **54**:890–897.

Haydon, D., R. R. Radtkey, and E. R. Pianka. 1993. Experimental biogeogaphy: interactions between stochastic, historical, and ecological processes in a model archipelago. Pages 108–116 *in* R. E. Ricklefs and D. Schluter, editors. Species diversity in ecological communities: historical and geographical perspectives. University of Chicago Press, Chicago, Illinois, USA.

Heck, K. L., Jr., G. van Belle, and D. S. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology **56**:1459–1461.

Herwitz, S. R., R. P. Wunderlin, and B. P. Hansen. 1996. Species turnover on a protected subtropical barrier island: a long-term study. Journal of Biogeography **23**:705–715.

Heywood, V. H., G. M. Mace, R. M. May, and S. N. Stuart. 1994. Uncertainties in extinction rates. Nature **368**:105.

Higgs, A. J., and M. B. Usher. 1980. Should nature reserves be large or small? Nature **285**:568–569.

Hines, J. E., T. Boulinier, J. D. Nichols, J. R. Sauer, and K. H. Pollock. 1999. COMDYN: software to study the dynamics of animal communities using a capture–recapture approach. Bird Study **46**:209–217.

Holt, R. D. 1992. A neglected facet of island biogeography: the role of internal spatial dynamics in area effects. Theoretical Population Biology **41**:354–371.

Jones, K. B., K. H. Riitters, J. D. Wickham, R. D. Tankersley, R. V. O'Neill, D. J. Chaloud, E. R. Smith, and A. C. Neale. 1997. An ecological assessment of the United States Mid-Atlantic Region. United States Environmental Protection Agency, Office of Research and Development, Washington, D.C., USA.

Kilburn, P. D. 1966. Analysis of species–area relations. Ecology **47**:831–843.

Kinzig, A. P., and J. Harte. 2000. Implications of endemic–area relationships for estimates of species extinctions. Ecology **81**:3305–3311.

Kobayashi, S. 1974. The species–area relation I. A model for discrete sampling. Researches on Population Ecology (Kyoto) **15**:223–237.

Kobayashi, S. 1975. The species–area relation II. A second model for continuous sampling. Researches on Population Ecology (Kyoto) **16**:265–280.

Kohn, D. D., and D. M. Walsh. 1994. Plant species richness—the effect of island size and habitat diversity. Journal of Ecology **82**:367–377.

Lack, D. 1973. The number of species of hummingbirds in the West Indies. Evolution **27**:326–337.

Leitner, W.-A., and M. L. Rosenzweig. 1997. Nested species–area curves and stochastic sampling: a new theory. Oikos **79**:503–512.

Lomolino, M. V. 1989. Interpretations and comparisons of constants in the species–area relationship: an additional caution. American Naturalist **133**:277–280.

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

Magsalay, P., T. Brooks, G. Dutson, and R. Timmins. 1995. Extinction and conservation on Cebu. Nature **373**:294.

Martin, T. E. 1982. Species–area slopes and coefficients: a caution on their interpretation. American Naturalist **118**: 823–837.

McGarigal, K., and B. J. Marks. 1994. FRAGSTATS. Spatial pattern analysis program for quantifying landscape structure, Version 2. [Available online.] URL:⟨http://www.fsl.orst.edu/lter/data/software/fragstat.htm⟩

McGuinness, K. A. 1984. Equations and explanations in the study of species–area curves. Biological Reviews of the Cambridge Philosophical Society **59**:423–440.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998a. Estimating rates of local extinction, colonization, and turnover in animal communities. Ecological Applications **8**:1213–1225.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998b. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology **6**:1390–1398.

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.

Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk **117**:393–408.

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–recapture data on closed animal populations. Wildlife Monographs **62**.

Palmer, M. W. 1990. The estimation of species richness by extrapolation. Ecology **71**:1195–1198.

Palmer, M. W. 1991. Estimating species richness: the second-order jackknife reconsidered. Ecology **72**:1195–1198.

Peterjohn, B. G., and J. R. Sauer. 1993. North American Breeding Bird Survey annual summary 1990–1991. Bird Populations **1**:52–67.

Pimm, S. L., G. J. Russell, J. L. Gittleman, and T. M. Brooks. 1995. The future of biodiversity. Science **269**:347.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models for mixtures. Biometrics **56**:434–442.

Pollock, K. H., J. E. Hines, and J. D. Nichols. 1984. The use of auxiliary variables in capture–recapture and removal experiments. Biometrics **40**:329–340.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. Wildlife Monographs **107**:1–97.

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.

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology **29**:255–282.

Preston, F. W. 1960. Time and space and the variation of species. Ecology **41**:611–627.

Preston, F. W. 1962a. The canonical distribution of commonness and rarity: Part I. Ecology **43**:185–215.

Preston, F. W. 1962b. The canonical distribution of commonness and rarity: Part II. Ecology **43**:410–432.

Preston, F. W. 1979. The invisible birds. Ecology **60**:451–454.

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.

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

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

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. American Naturalist **102**:243–282.

SAS Institute. 1988. SAS stat user's guide. SAS Institute, Cary, North Carolina, USA.

Simberloff, D. S. 1972. Properties of the rarefaction diversity measurement. American Naturalist **106**:414–418.

Simberloff, D. S. 1979. Rarefaction as a distribution-free method of expressing and estimating diversity. Pages 159–176 in J. F. Grassle, G. P. Patil, W. Smith and C. Taillie, editors. Ecological diversity in theory and practice. International Cooperative Publishing House, Fairland, Maryland, USA.

Stevens, G. C. 1986. Dissection of the species–area relationship among wood-boring insects and their host plants. American Naturalist **128**:35–46.

Walter, B. A., P. Cotgreave, R. D. Price, R. D. Gregory, and D. H. Clayton. 1995. Sampling effort and parasite species richness. Parasitology Today **11**:306–310.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA, in press.

Williams, C. B. 1943. Area and number of species. Nature **152**:264–278.

Wright, S. J. 1988. Patterns of abundance and the form of the species–area relation. American Naturalist **131**:401–411.