

ESTIMATING LANDSCAPE-SCALE SPECIES RICHNESS: RECONCILING FREQUENCY- AND TURNOVER-BASED APPROACHES

R. TODD JOBE¹

Ecology Curriculum, CB 3275, University of North Carolina, Chapel Hill, North Carolina 27599-3275 USA

Abstract. One hypothesis for why estimators of species richness tend to underestimate total richness is that they do not explicitly account for increases in species richness due to spatial or environmental turnover in species composition (beta diversity). I analyze the similarity of a data set of native trees in Great Smoky Mountains National Park, USA, and assess the robustness of these estimators against recently developed ones that incorporate turnover explicitly: the total species accumulation method (T-S) and a method based on the distance decay of similarity. I show that the T-S estimator can give reliable estimates of species richness, given an appropriate grouping of sites. The estimator based on distance decay of similarity performed poorly. There are two main reasons for this: sample size effects and the assumption that distance decay of similarity exhibits a power law relationship. I show that estimators based on distance–decay relationships exhibit systematically lower rates of distance decay for samples with few individuals per site independent of environmental variation. Second, the data presented here and many other survey data sets exhibit exponential rather than power law distance–decay relationships. Richness estimators that explicitly incorporate beta diversity can be improved by beginning from an exponential distance–decay relationship and adjusting for the systematic errors introduced by small sample sizes.

Key words: *beta diversity; distance decay of similarity; Great Smoky Mountains National Park, USA; species richness estimation.*

INTRODUCTION

The assessment of species richness over landscapes requires ever-increasing sample effort to capture the ever rarer species (Fisher et al. 1943, Preston 1948). Unless a landscape is completely surveyed, all samples fail to record a certain proportion of rare species. Consequently, in order to estimate total species richness (S_{true}) at a landscape scale (10^3 – 10^6 ha), we are forced to extrapolate from incomplete surveys of total richness. There is a long history in ecology of estimating total species richness (S_{est}) from sample data, and a wide variety of techniques for doing so (Chao 2004). Numerous papers have assessed the bias and precision of any of a number of estimators for a given taxonomic group and location (for reviews see Cao et al. 2004, Walther and Moore 2005).

Species richness estimators rely upon the relationship between species richness and the accumulation of sample effort or area to estimate the total number of species for some unmeasured amount of sample effort either as time, area, or number of individuals sampled. The relationship between species richness and sample effort is summarized as a species accumulation curve in which

the x -axis is increasing sample effort or number of individuals and the y -axis is increasing species richness.

Species richness estimators that use rarefied species accumulation curves fall into two broad categories: parametric and nonparametric. The former fit a function (typically a Michaelis-Menten) to the species accumulation curve (e.g., Plotkin et al. 2000, Jimenez-Valverde et al. 2006). The latter estimate species richness based on the frequency distribution of either species among sites (incidence based) or the number of individuals of each species (abundance based). These nonparametric estimators have more accurate estimates at small sample sizes than parametric ones and are typically preferred (Colwell and Coddington 1994).

An implicit assumption of nonparametric estimators is that sites are spatially homogeneous (Chazdon et al. 1998) and that the population from which individuals or species are drawn is stationary. As such, these estimators are typically used to estimate alpha diversity (Whittaker 1972), which is local species richness in which environment and other factors that control the species distributions are relatively constant. Recent evidence suggests, however, that these estimators are robust to spatial heterogeneity as long as sample coverage, the proportion of species observed relative to total species, is high (Brose et al. 2003, O’Dea et al. 2006). In fact, many studies that assess the performance (as measured by bias, precision, and/or accuracy) of these richness estimators were based on samples that could be

Manuscript received 11 October 2006; revised 8 May 2007; accepted 25 May 2007. Corresponding Editor: D. P. C. Peters.

¹ Present Address: Geography Department, CB 3220, University of North Carolina, Chapel Hill, North Carolina 27599-3220 USA.

E-mail: toddjobe@unc.edu

considered heterogeneous (e.g., Palmer 1990, Chiarucci et al. 2001).

In spite of their popularity, available nonparametric estimators typically underestimate species richness (Chao 1984, Colwell and Coddington 1994), possibly because they do not explicitly account for increases in species richness due to turnover in species composition across gradients. Such turnover, or beta diversity (Whittaker 1972), is responsible for greater species richness of large areas than would be suggested by extrapolating from small areas and is the driving force behind changes of richness with scale (Preston 1962a, b, Rosenzweig 1995). Beta diversity may therefore hold the key to understanding the manner in which richness scales from small areas that can be completely surveyed to large areas that cannot. Two recently published richness estimators (Harte et al. 1999, Ugland et al. 2003) explicitly use beta diversity in calculating S_{est} . The T-S estimator of Ugland et al. (2003) relies on grouping sites into ecologically meaningful subsets and integrating species richness estimates across different combinations of these groupings. The estimator developed by Harte et al. (1999) builds upon the log–log relationship between richness and area (Arrhenius 1921) to calculate S_{est} based on the distance decay of compositional similarity (Nekola and White 1999). This method has received attention in the literature for estimating the richness of microorganisms (Green et al. 2004, Horner-Devine et al. 2004) and landscape-scale vegetation (Krishnamani et al. 2004).

Here I assess the robustness of species richness estimators that explicitly incorporate species turnover relative to other, more popular estimators that do not. I analyze the relationships between the similarity of environments, geographic locations, and species composition. I demonstrate the sensitivity of turnover-based estimators to sample size and sample coverage. Extending this concept with a simulated example, I show that estimators based on distance–decay relationships exhibit systematically lower rates of distance decay for samples with few individuals per site independent of environmental variation. Finally, I show how estimators based explicitly on turnover and those based on the frequency distribution of richness among sites can be used in concert to accurately estimate landscape-scale species richness.

METHODS

Study site

The study uses data from the Great Smoky Mountains National Park (GSMNP, Tennessee–North Carolina, USA; 35°35' N, 83°33' W; Fig. 1) to illustrate and evaluate methods for estimating total species richness. The Park area is a little over 2000 km² (39 × 87 km); 95% of the park is forested. Forests range from high-elevation red spruce–Fraser fir (*Picea rubens*–*Abies fraseri*) and northern hardwood forest dominated by red maple (*Acer saccharum*), American beech (*Fagus*

grandifolia), and yellow birch (*Betula alleghaniensis*) to eastern hemlock (*Tsuga canadensis*) and pine–oak (*Pinus* spp.–*Quercus* spp.) forests on mesic and dry sites, respectively. At lower elevations, rich cove forests dominated by tulip poplar (*Liriodendron tulipifera*), American basswood (*Tilia americana* var. *heterophylla*), and red maple (*Acer rubrum* var. *rubrum*) are present. Within the Park, 129 native tree species (S_{true}) have been documented (165 including exotic species) (White 1982; P. White, unpublished data; P. White, J. Fridley, and J. Rock, unpublished data). Nomenclature follows Weakley (2006).

Data

The data set used for this analysis is a compilation of vegetation studies conducted in GSMNP spanning 1976–2004 (Fig. 1). Though each study in the compilation had its own research questions, they all record the presence of vascular plants in an area of 1000 m². I have limited this analysis to trees, because the actual number of species in the Park ($S_{true} = 129$) is known to within a few species. I have further limited the species list to only native trees because of the rapidly changing richness of exotic species in the flora. In the data set, 103 native tree species had recorded observations; the mean plot richness was 14.7 species/plot with a maximum of 34 species/plot. The final data set consisted of 805 plots after removing those lacking native trees.

Analysis

I generated incidence-based species accumulation curves (S_{obs}) (Colwell et al. 2004) for the Park and parametric and nonparametric species richness estimates using the software package EstimateS (Colwell 2005). There are many incidence-based richness estimators, so I limited this analysis to those that have been reported in the literature to perform best. Among the parametric equations, I used a fitted Michaelis-Menten (M-M) (Raaijmakers 1987, Colwell et al. 2004). The nonparametric estimators were the incidence coverage estimator (ICE) (Chazdon et al. 1998, Chao et al. 2000), Chao's incidence-based estimator ($Chao_2$) (Chao 1984, 1987), and the second-order jackknife estimator ($Jack_2$) (Burnham and Overton 1978, 1979, Smith and Vanbelle 1984, Palmer 1991). All these estimators were calculated based on rarefied species accumulation curves (S_{obs}) (Colwell et al. 2004).

Four different measures of site similarity were generated for the set of plots. For all pairs of sites, I calculated the Jaccard and Sorenson similarity of species composition and the Euclidean distance of normalized environment. The environmental variables included in the similarity analysis were elevation, hill shade (azimuth 135, altitude 45), and relative wetness (as measured by the topographic convergence index; Moore et al. 1991, Wolock and McCabe 1995, Yeakley et al. 1998). These variables taken together correspond to the important ecological gradients of energy flux, tempera-

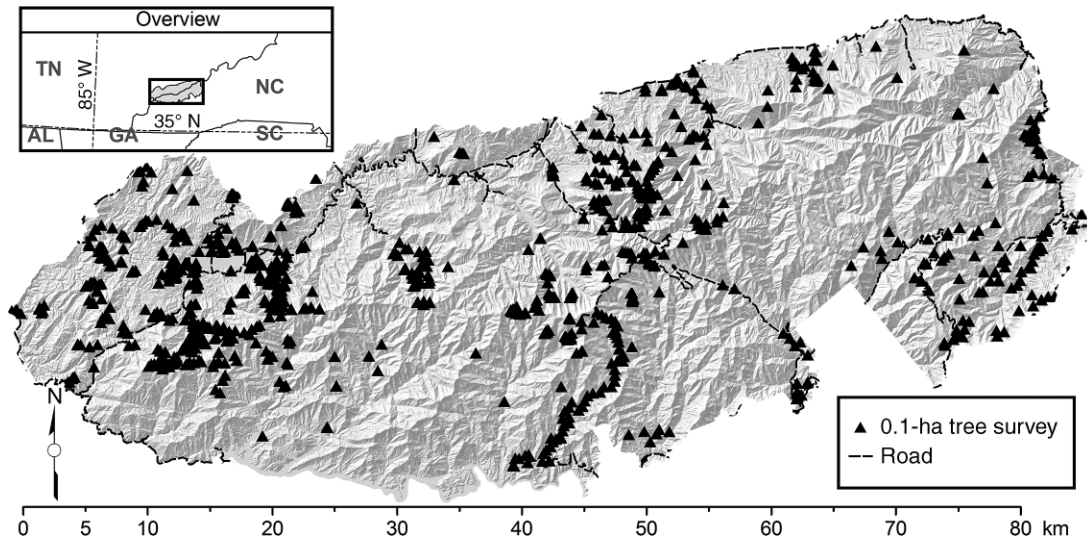


FIG. 1. Map of Great Smoky Mountains National Park (GSMNP), Tennessee–North Carolina, USA, showing the locations of the 805 tree survey plots used to estimate tree richness for the park.

ture, and radiation (Jobe 2006; R. T. Jobe, D. L. Urban, and P. S. White, *unpublished manuscript*). Since Bray-Curtis similarity is a rank order measure, each environmental variable had equal weighting. It is unlikely that species composition varies linearly with these environmental variables or that each variable is of equal importance for distance decay of similarity. The purpose of this environmental distance metric, however, was not to determine the functional relationship between environment and composition, but rather to cleanly describe environmental variation without resorting to an a priori functional relationship or one derived from the compositional data itself.

The two turnover-based estimators used in this analysis are the T-S estimator (Ugland et al. 2003) and the method of Harte et al. (1999). The T-S estimator relies upon groupings of similar sites. Given n groups, mean species accumulation curves are generated for all combinations of 1, 2, ..., n groups. Each combination has a mean maximum richness. These maximum values are then fit to a log linear species–area model. From this equation the total richness for a given area is calculated. I generated 10 groups three different ways: by species composition, by environment, and by geographic distance. I used the method of partitioning around medoids (PAM) (Kaufman and Rousseeuw 1990), a more robust version of k -means clustering, to assign group membership for each grouping variable.

The method of Harte et al. (1999) relies on the distance decay of compositional similarity (DDS) (Nekola and White 1999) to estimate species richness. The theory behind DDS builds upon Harte and Kinzig (1997). Beginning with the Arrhenius (1921) power law species–area relationship $S = cA^z$, where S is the number of species, A is area, and z is the slope of the log–log relationship, they derive the hypothesis that z is related

to the slope of a log–log distance decay of similarity (Sorenson similarity) by the function $z = -2d$, where d is the slope of the log–log distance decay. The slope of the log–log species–area relationship (z) is scale dependent (Rosenzweig 1995), so the method of Harte et al. (1999) is only applicable across scales in which z is constant. Harte has gone beyond this formulation to a model (HEAP) that does not assume a constant z across all scales (Krishnamani et al. 2004), but has yet to formulate a species richness estimator based on this model.

Analyzing the effect of sample size on the bias of the DDS estimator is not as straightforward as that for other richness estimators. Any random subset of sites can have a unique geographic extent. The DDS estimator relies on samples whose extent is at least as great as the square root of the area to which the extrapolation is made. To correct for this I generated smaller samples by first selecting a pair of plots randomly whose interplot distance was at least 40 km (roughly the square root of the park area). Additional sites were added randomly up to the desired sample size. This gave a random subset whose extent was fixed. The DDS estimator for the park area could be calculated on these subsets.

RESULTS

Similarity in species composition decreased with increasing distance between sites. Since distance measures are strongly influenced by edge effects (Nekola and White 1999), the smallest linear extent of the Park (the north-south extent) set the maximum distance for comparison among sites (40 km; Fig. 2a). This extent is somewhat larger than the ideal extent for reducing edge effects (one-half the minimum extent; Palmer 1988), but still reduces edge effects considerably due to

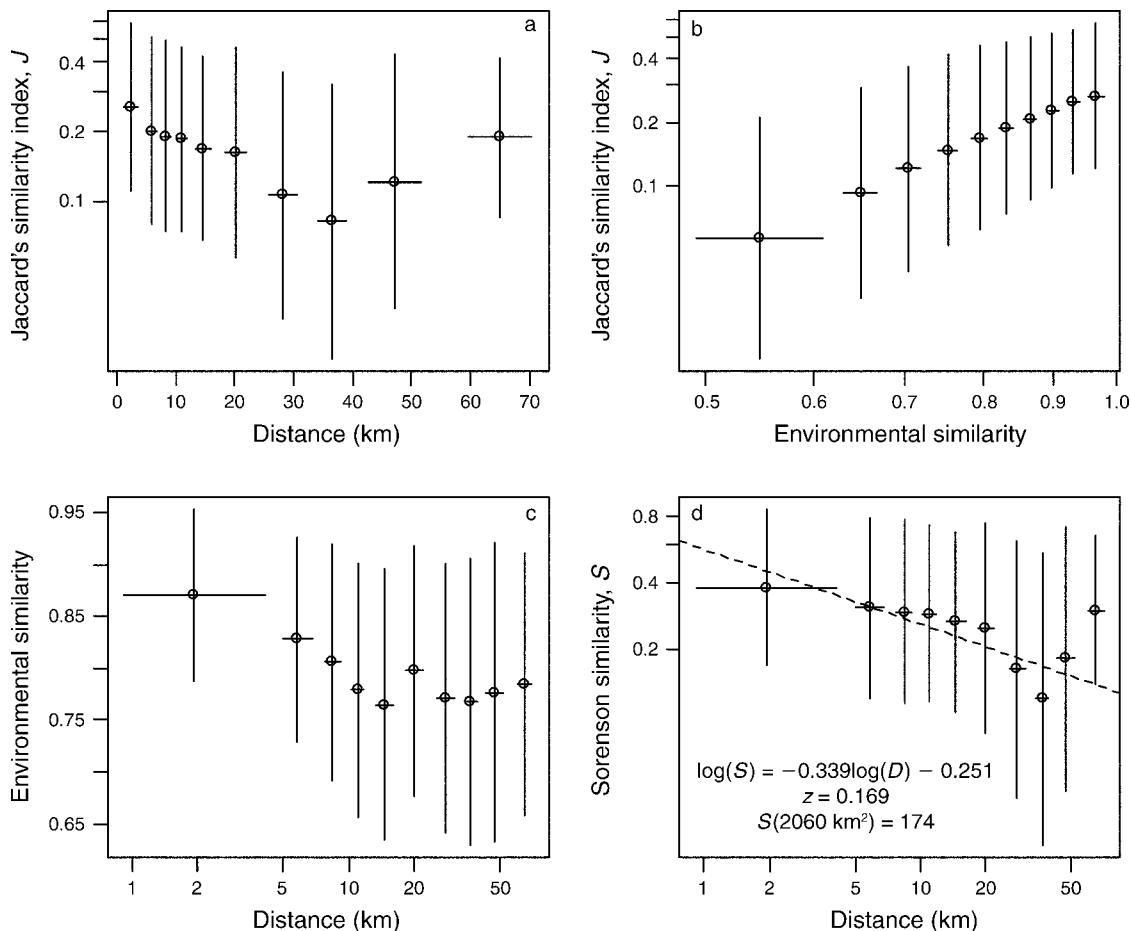


FIG. 2. Distance decay of similarity (mean \pm SD) for species composition and environment for 1000-m² vegetation plots in Great Smoky Mountains National Park. Circles are the mean value for each of 10 equal-sized groups of distances along the abscissa. Comparisons shown are: (a) compositional similarity (Jaccard's index, J ; log scale) by linear distance; this is the standard distance decay of similarity plot (sensu Nekola and White 1999); (b) compositional similarity (log scale) by environmental similarity (Bray distance; log scale); (c) linear environmental similarity vs. linear distance (log scale); and (d) Sorenson similarity (S ; log scale) vs. linear distance (D ; log scale), whose linearly regressed slope is equal to $-2z$ where z is the exponent of the Arrhenius (1921) species-area function (sensu Harte et al. 1999).

the elongated shape of the park. Similarity in species composition shows a more direct correlation with environmental similarity than distance (Fig. 2b). The relationship between environment and compositional similarity seems to be explained well by a log compositional similarity and linear environmental similarity. Distance and environment show a log linear relationship up to distances of ~ 20 km, at which point environment and distance seem uncorrelated (Fig. 2c).

The DDS estimator can only be applied to those distances < 40 km because of the edge effects described in the preceding paragraph (Fig. 2d). Since the Park area is roughly $45 \times 45 \text{ km}^2$, though, the DDS method can still be applied to estimate richness for the entire park. Harte et al. (1999) suggest a correction for rectangular areas that involves increasing the value of z . Since S_{est} from the DDS method was actually much larger than S_{true} , this correction was not applied.

The residuals of regressing $\log(\text{Sorenson similarity})$ against $\log(\text{geographic distance})$ to 40 km suggest that errors are not independent and thus violate one of the important assumptions of linear regression (Fig. 2d). This also corroborates other evidence suggesting that the distance decay of compositional similarity is log linear as opposed to log-log (Nekola and White 1999).

The classic parametric and nonparametric estimators of species richness underestimated native tree species richness in GSMNP by $\sim 20\%$ on average (Table 1). The similarity-based estimators performed better or overestimated species richness. Contrary to other results in the literature (Ugland et al. 2003, O'Dea et al. 2006) the T-S estimators performed the best of all the estimators. The DDS estimator was actually the poorest performer of all the estimators, overestimating species richness by 35%. The overestimation of species richness was actually

TABLE 1. Richness (S) estimates and performance for a variety of estimators.

Estimator	S_{est}	$S_{\text{est}}/S_{\text{true}}$	Model
S_{obs}	103	0.80	
Parametric			
Michaelis-Menten (M-M)	97	0.75	
Nonparametric			
Incidence coverage estimator (ICE)	105	0.81	
Chao ₂	104	0.81	
Jack ₂			
Similarity based			
Total species accumulation (T-S)			
Environmental distance	121	0.94	$14.3\ln(A)+15.7$
Geographic distance	125	0.97	$12.8\ln(A)+24.2$
Compositional similarity	122	0.95	$13.7\ln(A)+16.5$
Harte et al. (1999)	174	1.35	$(14.7)(A/0.001)^{0.17}$

Notes: See *Methods: Analysis* for explanations of estimators. S_{true} is 129 species; A is area. Contrary to previous studies, the total species accumulation method (T-S) estimators outperformed any other estimator. The method of Harte et al. (1999) was the poorest performer.

worse at small sample sizes, the opposite of all other estimators (Fig. 3).

DISCUSSION

Similarity in species composition, environment, and location

Environment tends to dominate changes in species composition at small scales (Nekola and White 1999), and this general pattern holds for my data to ~ 20 km. Beyond 20 km, species composition remains correlated with distance beyond that which environment explains, suggesting that either there exists important, yet unmeasured, environmental variables that are spatially autocorrelated at distances greater than 20 km or the signal of dispersal limitation in trees is present beyond 20 km. It is more likely the former, since I have not here considered disturbance history. The logging history within the park is likely to correlate with elevation, but may be structured at larger spatial grain than the combined environmental gradients of temperature, insolation, and relative wetness. For the purposes of species richness estimation, it is not necessary to tease apart the relative contribution of measured and unmeasured environmental gradients, nor their correlation with distance, but rather to show that turnover does occur and that it varies along the selected gradients.

This breakdown of the selected environmental gradients with distance also illustrates one reason why the species–area relationship deviates from a power–law relationship at both large scales and small scales (Rosenzweig 1995). That is, the slope of the log–log species area relationship (z) varies at both large scales and small scales. If the derivation of Harte et al. (1999) is correct, z is not constant from $1 \times 10^3 \text{ m}^2$ to $2 \times 10^5 \text{ m}^2$.

This limitation is overcome in practice by successively integrating over small changes in area, where changes in z are small (Hortal et al. 2006). Species estimates derived from extrapolation between the plot and some larger area (smaller than the landscape) are used in the calculation of S_{true} . This procedure has two shortcomings. First, since the parameter being estimated (z) is exponentially related to species number, small errors in estimating z yield drastic errors in estimating species number. Second, these errors in the estimation of z are multiplicative and are propagated when applied sequentially from small areas to large areas. As a result, the cumulative error in S_{est} is much larger than that for any single extrapolation of S_{obs} at a given scale.

Turnover-based estimators

The classic estimators of species richness (M-M, ICE, Chao₂, Jack₂, and their abundance-based counterparts) have their origin in methods for extrapolating true population size from mark–recapture sampling (Chao 1984). These estimators attempt to estimate the number of unobserved species in an unknown stationary population (of species). As sample coverage (the proportion of the entire pool of species observed in the sample) increases, the accuracy of the richness estimator increases. Brose et al. (2003) have shown that these metrics are relatively insensitive to environmental heterogeneity and spatial autocorrelation, so should perform well with samples that include a lot of heterogeneity and spatial autocorrelation, but they are relatively sensitive to sample coverage. In practice, though, increasing sample heterogeneity by adding new sites consummately increases the species pool. Sample coverage is also decreased because the species pool grows faster than the proportion of species captured in the sample. So, while gradients may not affect estimator performance directly, they affect their performance indirectly by making the universe of species bigger as dissimilar sites are added (Ugland and Gray 2004).

Turnover-based estimators are plagued by the same problems, but in a different way. The T-S estimator performed better than expected based on the results of other studies. In previous studies, the estimator always overestimated S_{true} by a substantial amount. The errors in species estimation are mainly due to the fact that choosing the number of groups and the membership in each group is somewhat arbitrary. Group membership in particular is important because all of the members are assumed to have the same species pool. In previous studies, group membership was decided based on either making equal-interval divisions across an ordination axis, an environmental gradient, or categorical habitat types. None of these methods asks the data which groupings are appropriate. Assigning group membership by nonhierarchical clustering (such as PAM) allows natural groupings of similar sites based on the data set. This is the likely reason for the better performance of this estimator, even when using a normalized environ-

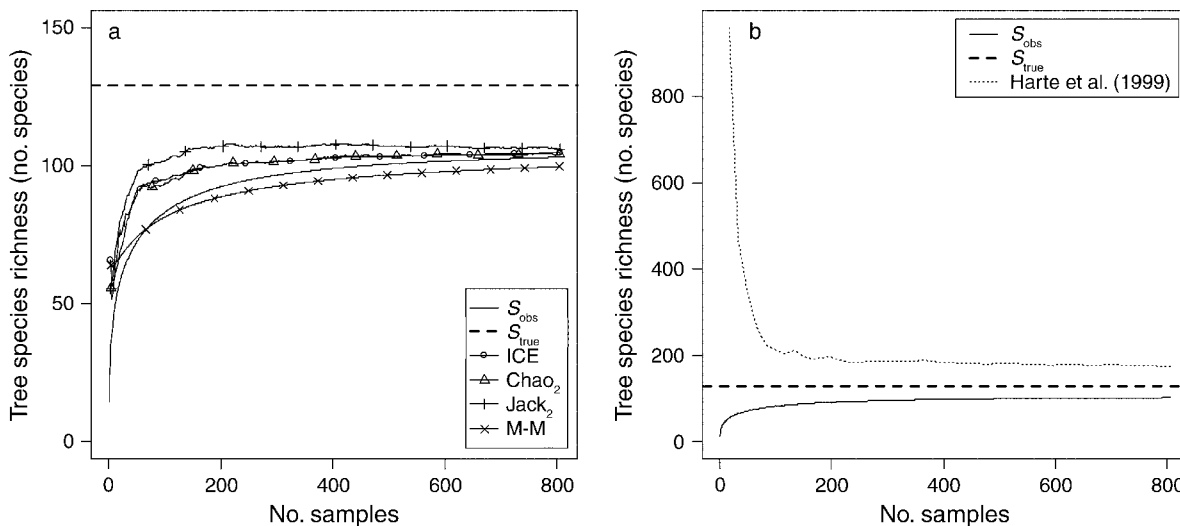


FIG. 3. Estimated species richness of trees in Great Smoky Mountains National Park with increasing sample size (no. 1000-m² vegetation plots) for (a) parametric and nonparametric estimators and (b) the turnover-based estimator of Harte et al. (1999). All estimators decrease bias with sample size, but the turnover-based estimator tends to overestimate richness at small sample sizes, while other estimators underestimate richness. For an explanation of abbreviations, see *Methods: Analysis*.

mental index that is probably nonlinear with respect to species composition. O'Dea and others (2006) suggest that the T-S estimator is unnecessary because the species–area relationship is implicit in estimators of species richness. Nevertheless, richness estimators not based on turnover always underestimate richness due to turnover between sites. The key to incorporating compositional turnover explicitly in species richness estimation lies in separating the differences in species composition between sites that are due to environmental turnover from those that are influences because sample coverage is too small.

The impact of sampling constraints on similarity

The increase of species with area beginning from the smallest scales and moving upward is a function of two processes. The first is ecology, that is, the sum total of dispersal limitation, environmental heterogeneity, and competition. The second is sampling constraints. That is, richness at small scale is constrained by the number of individuals that can fit in a given area (Fisher et al. 1943). As area increases, the dominance of sampling constraints becomes less and the ecological forces become greater. Since both ecological and sampling processes covary with grain, increasing sample grain is not equivalent to increasing sample size, especially for plants. The more individuals that are sampled, the more environmental heterogeneity is present and the greater the species pool. The solution proposed by Harte et al. (1999) is to increase the sample grain until the log–log relationship of species and area is constant. This area is likely to be quite large for trees (much greater than 1 ha), though for smaller organisms sample area is not as constraining (e.g., Green et al. 2004, Horner-Devine et al. 2004).

Similarity measures are also sensitive to sample coverage. Sample coverage is itself constrained by sample size. Smaller samples will systematically exhibit lower similarities than populations that have large sample sizes. Thus similarity measured in small samples is a biased estimator of the true similarity of two sites. More importantly, this bias is more pronounced for sites in which true similarity is high (Fig. 4a). As an example, consider two sites of 10 species each. Each site displays complete evenness, so the selection probability of species is equal. They have all species in common, so their actual Sorenson similarity is 1. The mean Sorenson similarity of many random draws of one individual from each site would be 0.1 because the probability of drawing two individuals of the same species is 1/10. This occurs purely because sample size is too low. Now consider the opposite case in which no species overlap between the two sites. The mean Sorenson similarity for many random draws of a single individual will be the true similarity between sites: 0. Sites with high similarity and low sample size exhibit greater bias toward low similarities than sites whose similarity is actually low (Fig. 4a).

The fact that estimates of similarity based on small samples show greater bias for sites of high similarity than low similarity has an important implication for the distance decay of similarity relationship (Fig. 4b). Since sites that are similar are more likely to be underestimated than sites that are very dissimilar, the effect of increasing the numbers of individuals per site would be to actually increase the rate of distance decay. As numbers of individuals per site increased, bias would decrease and similarity would increase. This increase would be greater for neighboring sites whose similarities are high than for distant sites whose similarities are low.

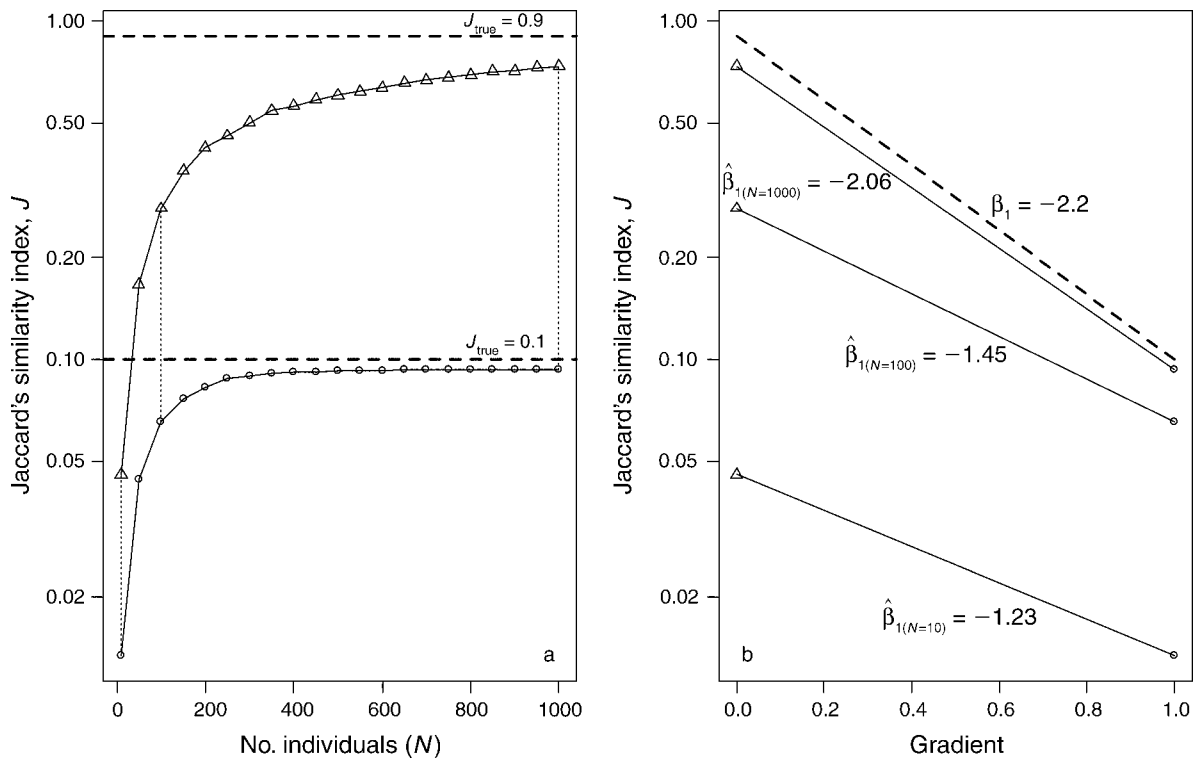


FIG. 4. Simulation results showing how sample size influences distance decay of similarity for a region with similar richness to Great Smoky Mountains National Park trees (129 species). (a) Results from two pairs of sites (low similarity, circles; high similarity, triangles) sampled at different intensities are compared with the actual similarities (J_{true} , dashed lines). (b) Results from three sample sizes (solid lines, $N = 10, 100, 1000$) along a hypothetical gradient. As the number of individuals per sample increases, similarity increases faster for similar sites (triangles) than dissimilar ones (circles). The actual change in similarity between the two communities (dashed line, β_1) is greater than that derived from any small sample (solid lines, $\hat{\beta}_{1,N}$).

If the area of sites were increased, the slope of distance decay of similarity would also increase. Thus, if sample size increases, then the DDS estimator should actually become worse. This is not necessarily the case, however, because increases in sample size at a particular location necessitate an increase in the environmental heterogeneity of the site, especially for plants. As discussed above, increasing sample size can actually decrease sample coverage because the number of species that could occupy a site, all else being equal, increases faster than the rate at which species are captured by the sample. As sample coverage decreases with increasing area, the similarity bias associated with small sample sizes returns. One solution to this problem might be to sum species numbers for a site through time (Adler et al. 2005, Fridley et al. 2006, White et al. 2006) but sites through time are subject to the same assumption of stationarity as sites through space, namely, disturbance or the shifting mosaic of landscape patches can cause the species pools for any given site to change through time. Below I describe an alternative to understanding the distance decay of similarity relationship and similarity-based richness estimators that accounts for the small sample effect without increasing the species pool.

Incorporating turnover in species richness estimators

The alternative to incorporating turnover into richness estimators involves combining the approaches of the point estimators with turnover-based estimators. Point estimators of stationary populations need to be used at scales and locations for which they are appropriate (i.e., within relatively homogeneous sites). Employed in this way with abundance data, the true or asymptotic similarity between two sites can be estimated. Chao et al. (2005) have developed just such a series of asymptotic similarity estimators that are analogous to the ones currently used in distance decay of similarity analysis. This removes, or at least removes the estimated effect of, low sample size. Then, using these similarity estimates for each site, one could apply the approach of Harte et al. (1999), which relates the distance decay of similarity to the accumulation of species with area. Unfortunately, that would only increase overestimation of S_{true} because the distance decay of similarity relationships would become steeper. More likely, the assumptions of power-law relationships between species, area, and distance decay of similarity are flawed. The exponential distribution tends to fit the observed distance decay of similarity relationships better than do power-law relationships across a wide array of

ecosystems and taxa (Nekola and White 1999; J. C. Nekola, *personal communication*), and the results presented here support that hypothesis. Exponential approaches relating species accumulation to distance decay of similarity and from there to power law species–area relationships (or even variable functional forms, e.g., Krishnamani et al. 2004) may be fruitful avenues for exploring new estimators, but derivations from first principles are not yet available.

It is important to note that at larger extents than those I have considered here, historic effects and dispersal limitation play an important role in compositional turnover, one that is separate from both environment and sampling considerations (Cody 1975). Where these biogeographic effects dominate, estimators that rely upon environmental turnover or power law species–area relationships will not predict species richness. At regional to continental scales species richness estimators that explicitly incorporate compositional turnover must address these biogeographic processes.

Conclusions

My results suggest that estimators that incorporate compositional turnover can provide reasonable estimates of species richness. Estimators that separate sampling processes from ecological ones offer the most potential for advances in estimating species richness, since estimators that do not explicitly include ecological processes consistently tend to underestimate species number. Further empirical and theoretical studies are needed to shed light on the interactions between similarity, richness, and sampling processes.

ACKNOWLEDGMENTS

This research was conducted while the author was under support of the W. C. Coker Fellowship from the Biology Department of the University North Carolina at Chapel Hill. The vegetation survey database for Great Smoky Mountains National Park used in these analyses was compiled by Jason D. Fridley, Peter S. White, Robert K. Peet, Dean L. Urban, Michael A. Jenkins, Alan S. Weakley, and an anonymous reviewer provided comments on early drafts. I also thank Jeff Nekola for his very helpful and extensive comments on later drafts.

LITERATURE CITED

- Adler, P. B., E. P. White, W. K. Lauenroth, D. M. Kaufman, A. Rassweiler, and J. A. Rusak. 2005. Evidence for a general species–time–area relationship. *Ecology* 86:2032–2039.
- Arrhenius, O. 1921. Species and area. *Journal of Ecology* 9:95–99.
- Brose, U., N. D. Martinez, and R. J. Williams. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* 84:2364–2377.
- Burnham, K. P., and W. S. Overton. 1978. Estimation of 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.
- Cao, Y., D. P. Larsen, and D. White. 2004. Estimating regional species richness using a limited number of survey units. *Ecoscience* 11:23–35.
- Chao, A. 1984. Nonparametric-estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11: 265–270.
- Chao, A. 1987. Estimating the population-size for capture recapture data with unequal catchability. *Biometrics* 43:783–791.
- Chao, A. 2004. Species richness estimation. Pages 7909–7916 in N. Balakrishnan, C. B. Read, and B. Vidakovic, editors. *Encyclopedia of statistical sciences*. Wiley, New York, New York, USA.
- Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* 8:148–159.
- Chao, A., W. H. Hwang, Y. C. Chen, and C. Y. Kuo. 2000. Estimating the number of shared species in two communities. *Statistica Sinica* 10:227–246.
- Chazdon, R. L., R. K. Colwell, J. S. Denslow, and M. R. Guariguata. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. Pages 285–309 in F. Dallmeier and J. A. Comiskey, editors. *Forest biodiversity research, monitoring and modeling: conceptual background and Old World case studies*. Parthenon, Paris, France.
- Chiarucci, A., S. Maccherini, and V. De Dominicis. 2001. Evaluation and monitoring of the flora in a nature reserve by estimation methods. *Biological Conservation* 101:305–314.
- Cody, M. L. 1975. Towards a theory of continental species diversities: bird distributions over Mediterranean habitat gradients. Pages 214–257 in M. L. Cody and J. M. Diamond, editors. *Ecology and evolution of communities*. Harvard University Press, Cambridge, Massachusetts, USA.
- Colwell, R. K. 2005. EstimateS: Statistical estimation of species richness and shared species from samples. Version 7.5. User's guide and application. (<http://purl.oclc.org/estimates>)
- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B* 345:101–118.
- Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85:2717–2727.
- 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.
- Fridley, J. D., R. K. Peet, E. Van Der Maarel, and J. H. Willems. 2006. Integration of local and regional species–area relationships from space-time species accumulation. *American Naturalist* 168:133–143.
- Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie. 2004. Spatial scaling of microbial eukaryote diversity. *Nature* 432:747–750.
- Harte, J., and A. P. Kinzig. 1997. On the implications of species–area relationships for endemism, spatial turnover, and food web patterns. *Oikos* 80:417–427.
- Harte, J., S. McCarthy, K. Taylor, A. Kinzig, and M. L. Fischer. 1999. Estimating species–area relationships from plot to landscape scale using species spatial-turnover data. *Oikos* 86:45–54.
- Horner-Devine, M. C., M. Lage, J. B. Hughes, and B. J. M. Bohannan. 2004. A taxa–area relationship for bacteria. *Nature* 432:750–753.
- Hortal, J., P. A. V. Borges, and C. Gaspar. 2006. Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology* 75:274–287.
- Jimenez-Valverde, A., S. J. Mendoza, J. M. Cano, and M. L. Munguia. 2006. Comparing relative model fit of several species-accumulation functions to local Papilionoidea and Hesperioidea butterfly inventories of mediterranean habitats. *Biodiversity and Conservation* 15:177–190.

- Jobe, R. T. 2006. Biodiversity and scale: Determinants of species richness in Great Smoky Mountains National Park. Ph.D. Dissertation. University of North Carolina, Chapel Hill, North Carolina, USA.
- Kaufman, L., and P. J. Rousseeuw. 1990. Finding groups in data: an introduction to cluster analysis. John Wiley and Sons, New York, New York, USA.
- Krishnamani, R., A. Kumar, and J. Harte. 2004. Estimating species richness at large spatial scales using data from small discrete plots. *Ecography* 27:637–642.
- Moore, I. D., R. B. Grayson, and A. R. Ladson. 1991. Digital terrain modeling—a review of hydrological, geomorphological, and biological applications. *Hydrological Processes* 5:3–30.
- Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26:867–878.
- O'Dea, N., R. J. Whittaker, and K. I. Ugland. 2006. Using spatial heterogeneity to extrapolate species richness: a new method tested on Ecuadorian cloud forest birds. *Journal of Applied Ecology* 43:189–198.
- Palmer, M. W. 1988. Fractal geometry: a tool for describing spatial patterns of plant communities. *Vegetatio* 75:91–102.
- 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* 72:1512–1513.
- Plotkin, J. B., et al. 2000. Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences (USA)* 97:10850–10854.
- Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* 29:254–283.
- 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.
- Raaijmakers, J. G. W. 1987. Statistical-analysis of the Michaelis-Menten equation. *Biometrics* 43:793–803.
- Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge University Press, Cambridge, UK.
- Smith, E. P., and G. Vanbelle. 1984. Nonparametric-estimation of species richness. *Biometrics* 40:119–129.
- Ugland, K. I., and J. S. Gray. 2004. Estimation of species richness: analysis of the methods developed by Chao and Karakassis. *Marine Ecology Progress Series* 284:1–8.
- Ugland, K. I., J. S. Gray, and K. E. Ellingsen. 2003. The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology* 72:888–897.
- Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28:815–829.
- Weakley, A. S. 2006. Flora of the Carolinas, Virginia, Georgia, and surrounding areas. University of North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, North Carolina, USA.
- White, E. P., P. B. Adler, W. K. Lauenroth, R. A. Gill, D. Greenberg, D. M. Kaufman, A. Rassweiler, J. A. Rusak, M. D. Smith, J. R. Steinbeck, R. B. Waide, and J. Yao. 2006. A comparison of the species–time relationship across ecosystems and taxonomic groups. *Oikos* 112:185–195.
- White, P. S. 1982. The flora of Great Smoky Mountains National Park: an annotated checklist of the vascular plants and a review of previous floristic work. U.S. Department of the Interior, National Park Service, Southeast Regional Office, Research Resource Management Report, Atlanta, Georgia, USA.
- Whittaker, R. H. 1972. Evolution and measurement of species diversity. *Taxon* 21:213–251.
- Wolock, D. M., and G. J. McCabe. 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in Topmodel. *Water Resources Research* 31:1315–1324.
- Yeakley, J. A., W. T. Swank, L. W. Swift, G. M. Hornberger, and H. H. Shugart. 1998. Soil moisture gradients and controls on a Southern Appalachian hillslope from drought through recharge. *Hydrology and Earth System Sciences* 2: 41–49.