Effects of Sample Standardization on Mean Species Detectabilities and Estimates of Relative Differences in Species Richness among Assemblages

Yong Cao,[1,*] Charles P. Hawkins,[1,†] David P. Larsen,[2,‡] and John Van Sickle[3,§]

1. Western Center for Monitoring and Assessment of Freshwater Ecosystems, Department of Watershed Sciences, Utah State University, Logan, Utah 84322;
2. Pacific States Marine Fisheries Commission, c/o U.S. Environmental Protection Agency (USEPA), Corvallis, Oregon 97333;
3. Western Ecology Division, National Health and Environmental Effects Research Laboratory–USEPA, Corvallis, Oregon 97333

Submitted December 7, 2006; Accepted April 2, 2007; Electronically published July 19, 2007

Online enhancements: appendixes, figures, tables.

abstract: Ecological surveys provide the basic information needed to estimate differences in species richness among assemblages. Comparable estimates of the differences in richness between assemblages require equal mean species detectabilities across assemblages. However, mean species detectabilities are often unknown, typically low, and potentially different from one assemblage to another. As a result, inferences regarding differences in species richness among assemblages can be biased. We evaluated how well three methods used to produce comparable estimates of species richness achieved equal mean species detectabilities across diverse assemblages: rarefaction, statistical estimators, and standardization of sampling effort on mean taxonomic similarity among replicate samples (MRS). We used simulated assemblages to mimic a wide range of species-occurrence distributions and species richness to compare the performance of these three methods. Inferences regarding differences in species richness based on rarefaction were highly biased when richness estimates were compared among assemblages with distinctly different species-occurrence distributions. Statistical estimators only marginally reduced this bias. Standardization on MRS yielded the most comparable estimates of differences in species richness. These findings have important implications for our understanding of species-richness patterns, inferences drawn from biological monitoring data, and planning for biodiversity conservation.

Keywords: species richness, mean species detectability, statistical estimators, species-occurrence distributions, sample representativeness, Lincoln-Petersen model.

A central aim of community ecology is to understand how and why species diversity changes over space and time (e.g., MacArthur 1965; Whittaker 1975; Rosenzweig 1995; Fridley et al. 2006). This knowledge is critical for effective resource management, biological monitoring, and biodiversity conservation (e.g., Gaston 1996; Hawkins et al. 2000). Ecological surveys provide the basic information regarding spatial patterns and temporal trends in species richness. Ideally, a survey would provide the complete list of species present for every assemblage of interest. However, limited resources dictate that we use small samples to characterize an assemblage, which we refer to as the collection of organisms of interest within a specified space. Such samples usually capture an unknown proportion of the species occurring in an assemblage. In other words, observed species richness is derived from data for which true species detectabilities are unknown.

Accurate estimates of the true species richness (TSR) of assemblages would be ideal, but unbiased estimates of the relative differences in species richness among assemblages (i.e., the ratio of species richness for a pair of assemblages) are often sufficient for detection of patterns of species richness and tests of ecological hypotheses. Just as relative differences in density can be compared only when the capture probabilities of individuals are constant across populations (e.g., Pollock et al. 2002; Bailey et al. 2004; MacKenzie et al. 2006), unbiased estimates of the relative

* Corresponding author; e-mail: yong.cao@usu.edu.

† E-mail: chuck.hawkins@usu.edu.

‡ E-mail: larsen.phil@epa.gov.

§ E-mail: vansickle.john@epa.gov.

Am. Nat. 2007. Vol. 170, pp. 381–395. 0003-0147/2007/17003-42262$15.00.
DOI: 10.1086/520117

difference in observed species richness are possible only if mean species detectabilities (MSD), that is, average detectabilities across all species in an assemblage, are equal across assemblages (e.g., MacKenzie and Kendall 2002; Kery and Schmid 2006). The importance of this requirement can be easily illustrated. If the numbers of species in two assemblages are N_i and N_j, the true relative difference is $\lambda = N_i/N_j$. Since N_i and N_j are unknown, we estimate their ratio $\lambda = N_i/N_j$ based on the numbers of species (n_i and n_j) observed in the two assemblages. The expected value of the observed number of species in an assemblage is $E[n] = dN$, where d is mean species detectability (MSD). Thus, a robust estimator of λ is given by $(n_i/d_i)/(n_j/d_j)$ (MacKenzie and Kendall 2002). Unless $d_i = d_j$, the naive ratio (n_i/n_j) would clearly give a biased estimate. However, it is not clear how well current methods used for estimating relative differences in species richness generally meet the requirement for equal MSDs.

Comparisons of the relative differences in species richness among assemblages have been based on estimates of species richness rarefied to a constant sampling area or number of individuals (e.g., Sanders 1968; Hurlbert 1971; Gotelli and Colwell 2001), statistical extrapolations to TSR (e.g., Bunge and Fitzpatrick 1993; Colwell and Coddington 1994; Walther and Moore 2005), or richness standardized on a measure of sample representativeness (Cao et al. 2002*a*, 2004). The degree to which these three methods achieve a constant MSD across assemblages will affect our perception of how species richness changes over space and time.

Rarefaction estimates the species richness expected for a given number of individuals or replicate samples that are randomly drawn from an assemblage (Gotelli and Colwell 2001). It is equivalent to standardizing sampling effort at the lowest effort used across assemblages. However, the shapes of species accumulation curves usually vary from one assemblage to another (e.g., Rosenzweig 1995; Witman et al. 2004). Rarefied samples containing the same number of individuals (individual-based rarefaction) or sampling units (sample-based rarefaction) may therefore reach different MSDs (e.g., Angermeier and Smogor 1995; Condit et al. 1996; Cao et al. 2002*a*; Watson 2003). In other words, samples standardized on either the number of individuals or the sample area/volume may reach different MSDs in different assemblages. As a result, estimates of relative differences in species richness among assemblages can be biased, but it is unclear how severe and general the bias is.

A second method uses statistical models to estimate TSR via extrapolation (e.g., Colwell and Coddington 1994). Those models were explicitly designed to address the fact that not all species are detected during a survey (e.g., Boulinier et al. 1998; Williams et al. 2001). Those models can be used to estimate relative differences in species richness only if they either accurately predict TSR or exhibit the same prediction bias across assemblages that differ in MSD; that is, the same proportion of TSR is predicted across assemblages. However, many studies have shown that thse models usually severely underestimate TSR unless sampling effort is extremely high (e.g., Hellmann and Fowler 1999; Wagner and Wildi 2002; Brose et al. 2003; Chiarucci et al. 2003; Cao et al. 2004; O'Dea et al. 2006). Many studies either explicitly or implicitly assume that statistical models can satisfactorily reduce bias when estimating relative differences in species richness (e.g., Winkler and Kampichler 2000; Melo 2004; Witman et al. 2004; Walther and Moore 2005); however, this assumption has never been tested.

A third method for estimating true differences in relative species richness among assemblages is based on the concept of sample representativeness, that is, how well a sample represents the species composition of the assemblage from which it is drawn (Cao et al. 2002*a*). When sample representativeness is measured with the average Jaccard coefficient (AJC; Jaccard 1901) calculated among multiple pairs of replicate samples, it estimates MSD. The Jaccard coefficient is calculated as the ratio of the number of shared species by two samples (S_{12}) to the total number of species recorded in the first (S_1) and second (S_2) samples; that is, $S_{12}/(S_1 + S_2 - S_{12})$. The Jaccard coefficient is mathematically equivalent to the Lincoln-Petersen model used in capture-recapture studies of individual populations. The latter measures mean individual detectability (d) as the ratio of the number of animals recaptured (m) to the number of animals marked (n); that is, $d = m/n$ (Williams et al. 2001). The two models are equivalent because $(S_1 + S_2 - S_{12}) = n$, $S_{12} = m$ (Cao et al. 2004). Cao et al. (2002*a*) compared species richness across sites on a standard level of MSD by repeatedly sampling an assemblage until a targeted value of AJC was reached. Because the Jaccard coefficient is typically employed to measure the taxonomic similarity between two assemblages (e.g., Legendre and Legendre 1998) and some concerns have been raised regarding its effectiveness for that specific purpose (e.g., Wolda 1981; Cao et al. 2002*b*; Chao et al. 2005), use of the term AJC in the context of species detectabilities could create unnecessary confusion in the literature. In the remainder of this article, therefore, we refer to estimates of MSD across multiple pairs of replicate samples as mean replicate similarity (MRS) as estimated by the Lincoln-Petersen model.

Although the accuracy of TSR estimates depends on how well estimates of MRS agree with MSD, the performance of the MRS standardizations to achieve similar MSDs across assemblages relies on how consistently MRS estimates are correlated with true MSDs across assemblages

(Cao et al. 2002*a*). The MRS is therefore used as an index of MSD in this standardization rather than a measure of MSD in estimating the true species richness. Studies have shown that the MRS-MSD relationship was highly consistent across different assemblages in several data sets (Cao et al. 2001, 2002*a*, 2004) that expressed MSD as %TSR. However, we do not know how general these relationships are because assemblages differ widely in species-abundance or species-occurrence distributions, TSR, and spatial patchiness within an assemblage. All of these factors can affect estimates of both species richness (e.g., He and Legendre 2002) and compositional similarity between replicate samples (e.g., Plotkin and Muller-Landau 2002).

Clearly, all three methods have potential limitations and uncertainty when used to estimate the relative differences in species richness between assemblages. A systematic evaluation of their performances across a wide range of assemblages is needed. Simulation provides an effective avenue for creating diverse assemblages with known TSR and unlimited numbers of replicates (e.g., He and Legendre 2002; Brose et al. 2003). In this article, we used simulated assemblage data to examine (1) how rarefied samples differ in MSD under a range of species-occurrence distributions, (2) how sensitive estimates of TSR derived from statistical models are to variation in MSD, (3) how the performance of the MRS-based standardization is affected by the difference in species-occurrence distributions and levels of TSR, and (4) under what conditions MSDs are most likely to differ in comparisons of species richness.

Data Sets and Simulation

Simulating assemblages typically requires the combined use of two types of models. The first type describes how individuals are distributed among species—that is, species-abundance distribution—or how species are distributed among samples—that is, species-occurrence distribution (SOD). Species-abundance distributions have been frequently reported to fit logseries or lognormal models (e.g., Fisher et al. 1943; Preston 1948; Williams 1964; May 1975; Gray 1978; Bell 2001; Hubbell 2001). The SODs are often described with patch-occupancy models (e.g., Hanski and Gyllenberg 1993; Gibson et al. 1999). The second type of models describes how individuals in a species are distributed within its habitat (i.e., spatial patchiness). Poisson and negative binomial distributions are commonly observed (Williams et al. 2001).

In our analysis, we simulated assemblages based on both species-abundance and SOD models, responding to individual-based and area-based sampling, respectively. However, we focused on the simulation of diverse SODs, which is essential for realistically mimicking a wide variety of assemblages. The results, however, should be applicable to logseries and lognormal species-abundance distributions because these distributions are special cases of SOD models under certain assumptions, as we show later. We also chose not to incorporate the possible effects of spatial patchiness in species distributions within an assemblage into the simulation at this time. This simplification was reasonable given that replicate samples used in practice (e.g., a 0.1-m^2 quadrat) are often substantially smaller than the area occupied by a targeted assemblage (e.g., a 100-km^2 lake), and random sampling used for estimating species richness and MRS can homogenize the effect of spatial patchiness on estimates of MRS and richness. Studies have shown that the spatial distributions of individuals are much less important than either species-abundance distributions or sampling effort in estimating species richness when similar sampling schemes were used (e.g., Brose et al. 2003; O'Dea et al. 2006).

Empirical Data

We used seven published data sets (25 total study areas or locations) for modeling SODs (app. A in the online edition of the *American Naturalist*). These data sets were derived from surveys conducted at different spatial scales and for different taxonomic groups (table A1). Species-level identification was used in all data sets except for those of macroinvertebrates, which were identified at mixed resolution, a common practice in freshwater macroinvertebrate studies because of the difficulty in identifying immature individuals (Carter and Resh 2001). These data sets are all based on extensive sampling; however, most of them surely did not contain all species occurring in the location. We refer to the number of species or taxa observed at a location as "TSR_{obs}" and use "TSR" for true species richness. We also use the term "sample" to mean a single replicate or a combination of replicates and the term "sampling effort" to mean the number of pooled replicates.

Modeling Species-Occurrence Distributions

The distribution of species among samples is often described with patch-occupancy models (e.g., Hanski and Gyllenberg 1993; Gibson et al. 1999). However, these models group species into arbitrary categories based on their occurrence frequencies (McGeoch and Gaston 2002) and assume that different species are randomly distributed within each of the occurrence frequency categories. These properties do not allow accurate description of species-occurrence probabilities, which are critical for estimating species richness and MRS. We therefore modeled the distributions differently. First, we estimated the probability

that species i occurred in a single randomly selected sample as $P_i = k_i/K$, where K is the total number of samples and k_i is the number of samples where species i occurs. The term P_i is also referred to as species occupancy (e.g., MacKenzie et al. 2006). When P_i values for all species are plotted against $X_i = \text{rank}(P_i)$ in descending order, the result is a decreasing curve whose shape expresses the pattern of decreasing commonness of species. We refer to this P_i, X_i curve as a species-occurrence distribution (SOD). When P_i is replaced by relative abundance, this curve simply becomes Whittaker's relative abundance–rank order plot (e.g., Ulrich and Ollik 2005).

The SODs observed in the seven empirical data sets (table A1) were adequately fit by one of three models: negative exponential (eq. [1]), logistic (eq. [2]), or linear (eq. [3]):

$$P_i = ae^{bX_i}, \tag{1}$$

$$P_i = \frac{1}{1 + ae^{bX_i}}, \tag{2}$$

$$P_i = a + bX_i, \tag{3}$$

where $X_i = 1, 2, \ldots,$ TSR. Parameter a controls the occurrence probability of the most common species, and parameter b controls the rate at which the occurrence probability of subsequent species decreases. Negative exponential models predict a few common species and a large proportion of rare species, and these types of assemblages are often species rich (table A1). Logistic models typically predict a large proportion of common species, but assemblages can be either species poor or species rich. Linear models predict occurrence probabilities that decrease at a constant rate, and these assemblages are usually species poor.

Defining Assemblages

We examined the ranges of parameter values for each of the types of fitted models and selected a set of values from which we created 12 assemblages. Each of them was defined by a specific SOD based on the model parameters (table 1). The ranges of TSR in assemblages simulated for each type of model were related to those observed in the empirical data. We set five levels of TSR (30–300 species) for the negative exponential model, four levels (15–300 species) for the logistic model, and three levels (25–100 species) for the linear model. We also specified 10 assemblages (TSR = 25–156) based on logseries and lognormal models (app. B in the online edition of the *American Naturalist*). These two models are a special case of SODs when individual-based sampling is used (table B1; fig. B1), and they were therefore not pursued further.

Table 1: Model parameters selected for 12 artificial assemblages (A1–A12) that follow three types of species-occurrence distributions

Species-occurrence distribution models	a	b	TSR
Negative exponential; $P_i = ae^{bXi}$:			
A1	1.25	−.250	30
A2	1.10	−.150	50
A3	1.00	−.100	75
A4	.90	−.075	100
A5	.70	−.025	300
Logistic; $P_i = 1/1 + ae^{bXi}$:			
A6	.0183	.50	15
A7	.0025	.30	30
A8	.0183	.04	200
A9	.1353	.02	300
Linear; $P_i = a + bX_i$:			
A10	1	−.04	25
A11	1	−.02	50
A12	1	−.01	100

Note: a = model parameter that controls the occurrence probability of the most common species; b = model parameter that controls the rate at which the occurrence probability of subsequent species decreases; TSR = true species richness.

Analytic Methods

Estimating Species Richness Based on Sample-Based Rarefaction and Statistical Estimators

Given a set of occurrence probabilities, P_i, $i = 1, \ldots,$ and k for k species, a single random replicate sample can be generated from a set of k Bernoulli trials, one for each species. We generated 2,000 random replicates for each of the 12 assemblages. We used EstimateS, version 7.5 (Colwell 2005), to calculate species richness for sample-based rarefaction. We then estimated TSR using multiple statistical methods available in the EstimateS 7.5 that were applicable to species presence-absence data, including Chao-1, Chao-2, incidence-based coverage estimator, first- and second-order jackknife, bootstrapping, and two variants of Michaelis-Menten models (app. C in the online edition of the *American Naturalist*). We examined the estimates of TSR by each of these estimators at 1–50 replicates, sampling efforts that are generally practical. The second-order jackknife method (Jack-2; Burnham and Overton 1978) was most accurate for 1–50 replicates, particularly for those assemblages exhibiting negative exponential SODs (see fig. C1 for an example). This result agreed with many other studies (e.g., Hellmann and Fowler 1999; Brose et al. 2003; Melo 2004, Hortal et al. 2006; but see Cormack 1989). We therefore used only the Jack-2 estimator, as-

suming that the other estimators tested would perform no better in estimating relative difference in species richness.

Establishing MRS-MSD Relationships

We followed Cao et al. (2002*a*) in using a random resampling procedure to establish MRS-MSD relationships. With N replicate samples collected from an assemblage, half are randomly drawn and pooled into a single sample. The other half are combined into the second sample. The similarity is calculated for this pair of samples, and the number of species in each of the two samples is counted. This process is repeated 1,000 times to obtain MRS and average species richness per sample, which was then converted to MSD.

A theoretical relationship between MRS and MSD can be modeled by progressively pooling a collection of independent replicates drawn at random, assuming a given SOD and TSR (app. D in the online edition of the *American Naturalist*). We constructed an MRS-MSD curve for each of the 12 assemblages.

Setting Evaluation Criteria

Accurate estimation of the relative differences in species richness among assemblages depends on how similar the estimates of MSDs are across assemblages. We therefore evaluated the accuracy of a method in estimating relative differences in richness by comparing the range and standard deviation (SD) of MSDs estimated across assemblages. We assessed how these two measures of variability differed within each type of SOD and across the three types of SODs. We compared the three methods (rarefaction, Jack-2, and standardization on MRS) in two ways.

We first compared the three methods with the range and SD of MSD among the 12 assemblages over a range of sampling efforts or MRS values. For MRS-based standardizations, both the specific values of MRS and their ranges differed across assemblages even though the sampling effort used was otherwise similar. It was therefore impossible to precisely compare the variability in MSD for exactly the same MRS values among different assemblages. We therefore fit the MRS-MSD curve for each assemblage and then calculated the range and SD in fitted MSD for specific values of MRS. These comparisons are indicative of the overall performances of the three methods. However, they cannot be compared directly because any specific level of MRS requires different sampling efforts in different assemblages; hence, MRS cannot be transformed to a consistent number of replicates across assemblages.

To overcome this difficulty, we took a second approach, which compared the three methods based on the same total sampling effort used for all assemblages (see app. E in the online edition of the *American Naturalist* for details). When q replicates are collected from each assemblage as required by rarefaction, the total sampling effort across the 12 assemblages is $12q$. The number of replicates required for a targeted MRS value differed among the assemblages (q_i for assemblage i); however, the total sampling effort (Σq_i) was constrained by $12q$. For the 12 artificial assemblages, the procedure was simplified. An MRS-based standardization requires a minimum of two replicate samples from an assemblage. The maximum MRS value reached by two replicates across the 12 assemblages sets the minimum level of MRS that can be used for standardization. This value was 0.6 (based on resampling the 2,000 replicates created earlier), for which two to 24 replicates were required, depending on which of the 12 assemblages was analyzed. The total effort across assemblages needed to achieve this value of MRS was 78 replicates (6.5 replicates per assemblage). We therefore calculated average species richness observed or extrapolated from seven replicates (rounded up from 6.5). We then compared the range and SD of MSDs among the 12 assemblages as calculated for each of the three methods. We recognize that such comparisons can show only which method will be most accurate in general and that the three methods might perform similarly under some situations.

Evaluating the Precision of MRS-MSD Relationships

The practical utility of species richness comparisons based on MRS-standardizations depends on the precision of MRS-MSD relationships, especially when comparisons are based on small numbers of replicates. We therefore need to know how the type of SOD, the level of TSR, and the value of MRS affect the precision of this relationship.

For the negative exponential model, we compared precision for a species-poor assemblage (A2, TSR = 50) and a species-rich assemblage (A5, TSR = 300). For the other two models, high MRS and MSD can be easily reached with low sampling effort when species richness is low. We therefore compared precision estimates obtained from the A8 (TSR = 200) and A9 assemblages (TSR = 300; logistic model) and precision estimates from the A11 (TSR = 50) and A12 (TSR = 100) assemblages (linear model).

The 2,000 replicates created earlier for establishing species curves were combined into at least 10 distinct subsets of four to six sampling efforts. The range of sampling efforts examined varied from one assemblage to another so that the ranges of MRS examined were as comparable as possible among the 12 assemblages: eight to 30 replicates for logistic and linear SOD assemblages and 20–200 replicates for negative exponential SOD assemblages. Each of the distinct subsets (q replicates per subset) was randomly split into two equal-sized samples (one-half q replicate

each), and assemblage similarity and species richness per one-half q replicates were calculated. This process was repeated 1,000 times to obtain estimates of MRS and MSD (see Cao et al. 2002*a*). We then plotted MSD against MRS across all subsets and at all sampling efforts examined for each assemblage. We estimated precision as the root mean squared error (RMSE) associated with the MRS-MSD curves derived from equations (4) and (5). We also correlated the RMSE with MRS values to assess whether precision was associated with the value of MRS used for standardization.

Results

Fitting Species-Occurrence Probability Models

Of the 25 different SODs examined (app. A), 14 were best described by negative exponential models (56% of cases), nine by logistic models (36% of cases), and two by linear models (8% of cases; fig. 1; table A2). The fit was adequate in all cases, with RMSE ranging between 0.0001 and 0.09. The stream macroinvertebrate assemblages at both local and regional scales, fish assemblages at the regional scale, and the tropical forest tree assemblage fit negative exponential models (fig. 1*A*–1*C*). The regional bird assemblages fit logistic models (fig. 1*D*). Local fish assemblages fell into all three types of SOD: negative exponential (36%), logistic (50%), and linear (14%) models (see fig. 1*E*, 1*F* for examples).

Comparing Species Richness Derived from Sample-Based Rarefaction

The MSD reached by a rarefied sample was much lower in assemblages exhibiting negative exponential SODs than in assemblages characterized by either logistic or linear SODs over the range of 1–50 pooled replicates (fig. 2*A*). For example, the average MSD based on a single replicate for the five negative exponential SOD assemblages was only 0.12, but it was 0.48–0.49 for the seven assemblages with either logistic or linear SODs. These differences did not significantly decrease with increasing sampling effort (at least up to 50 replicates). However, the differences in MSD among the four logistic SOD assemblages were also high (up to 0.3, SD = 0.12; fig. 2*B*). Forty replicates were required to reduce the difference in MSD among these assemblages to 0.05. In comparison, MSD varied much less among either the linear SOD (<0.01) or the negative exponential SODs (<0.07) assemblages (fig. 2*B*), which implies that rarefaction works well when compared assemblages all exhibit either linear or negative exponential SODs.

The varying levels of MSD reached by standardizing samples with rarefaction led to severe biases in estimating relative differences in species richness among assemblages when assemblages were sampled with low effort. For example, both A5 and A9 assemblages contained 300 species; however, the average number of species observed in one replicate from assemblage A9 (107 species) was 4.5 times higher than that observed in A5 (24 species) and 2.3 times higher (232 vs. 102 species) when 10 replicates were used. Furthermore, species accumulation curves frequently crossed one another with increasing sampling effort, indicating that even the rank of assemblages in observed species richness depended on the sampling effort used (fig. C2). For example, the observed richness in assemblage A12 was higher than in A5 until sampling effort exceeded seven replicates, after which observed richness was higher in A5.

Comparing Species Richness Based on Jack-2 Estimates

Jack-2 estimates always achieved higher proportions of TSR or MSD than raw samples (cf. figs. 2*A*, 3*A*). However, the differences in MSD among the three types of SODs remained largely because the predicted richness in the logistic and linear SOD assemblages approached plateaus much faster than did negative exponential SOD assemblages. The maximum difference in MSD between negative exponential and linear SOD assemblages ranged between 0.65 at two replicates and 0.30 at 50 replicates. The difference in MSD between negative exponential and linear SOD assemblages was similarly high, 0.26–0.60 over the same range of sampling effort. Jack-2 also increased the variation in MSD among the logistic SOD assemblages, although it slightly reduced the variation in MSD among the negative exponential SOD assemblages (fig. 3*B*) compared with rarefied samples (fig. 2*B*). In general, the Jack-2 estimator did little to reduce the bias in estimating relative differences in species richness that was observed for rarefied samples.

Comparing Richness Based on MRS

The MRS increased with increasing sampling effort in all 12 assemblages. The MRS sampling effort curves approximately fit power functions. However, the level of MRS reached at a given sampling effort differed greatly among assemblages with different types of species-occurrence distributions. The MRS was lowest for assemblages with negative exponential distributions and highest for those with logistic distributions. In other words, a standard sample reached very different mean species detectabilities in different assemblages. The MRS also decreased with increasing TSR within each type of species-occurrence distribution.

The MRS-MSD curves were adequately fit with three-

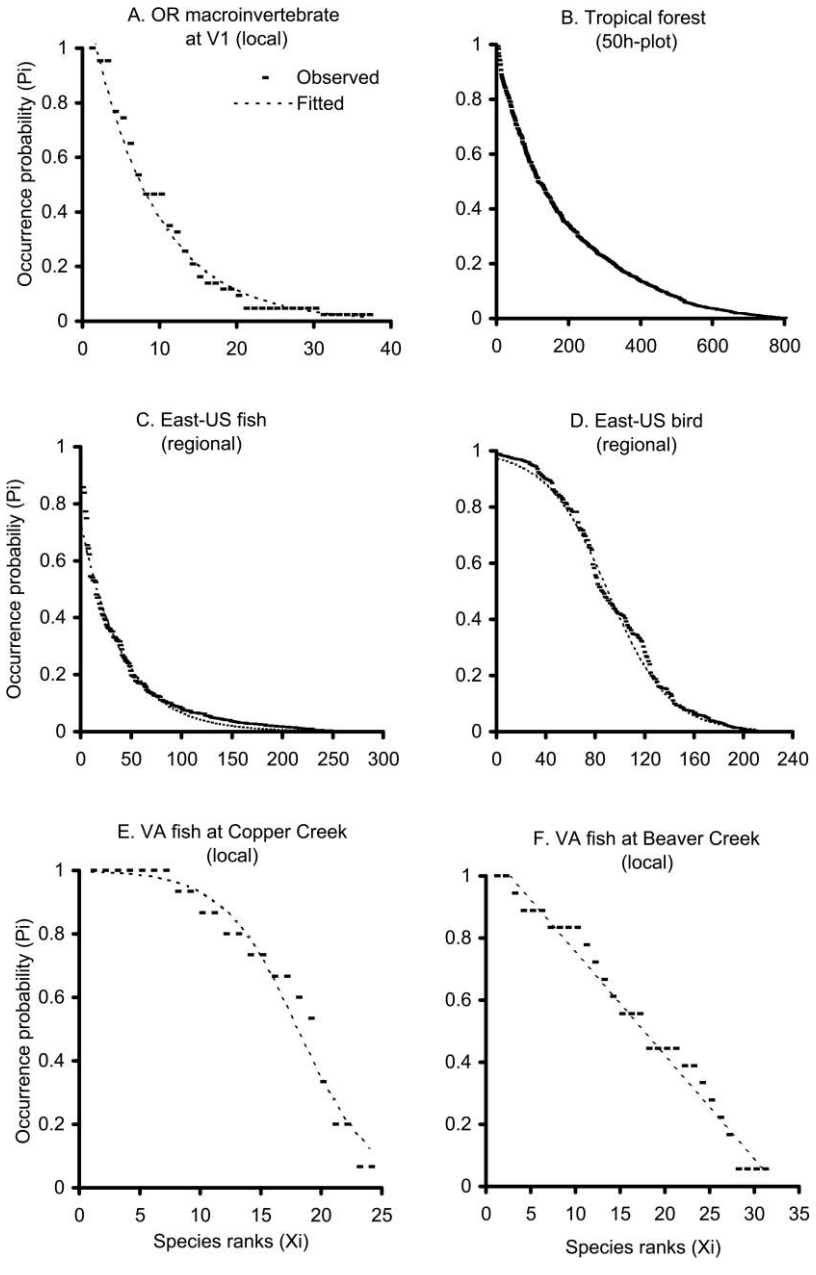

Figure 1: Species-occurrence distributions for four different taxonomic groups and at two spatial scales fit negative exponential (*A*–*C*), logistic (*D*–*E*), or linear models (*F*).

parameter logistic models (eq. [4]) for both negative exponential and logistic SOD assemblages and with linear models (eq. [5]) for linear SOD assemblages (table C1).

$$\text{MSD} = \frac{1}{a + e^{b+c\text{MRS}}}, \tag{4}$$

$$\text{MSD} = a + b \times \text{MRS}, \tag{5}$$

where parameters a and b set the predicted value of MSD when MRS is equal to 0 and 1, respectively. Parameter c controls the shape of the MRS-MSD curve.

The MRS-MSD relationships differed across the three types of SODs (fig. 4*A*) but to a much lesser extent than those species curves based on rarefied samples or predicted (Jack-2) richness (figs. 2*A*, 3*A*). For the same MRS value, MSD was usually the lowest for negative exponential SOD

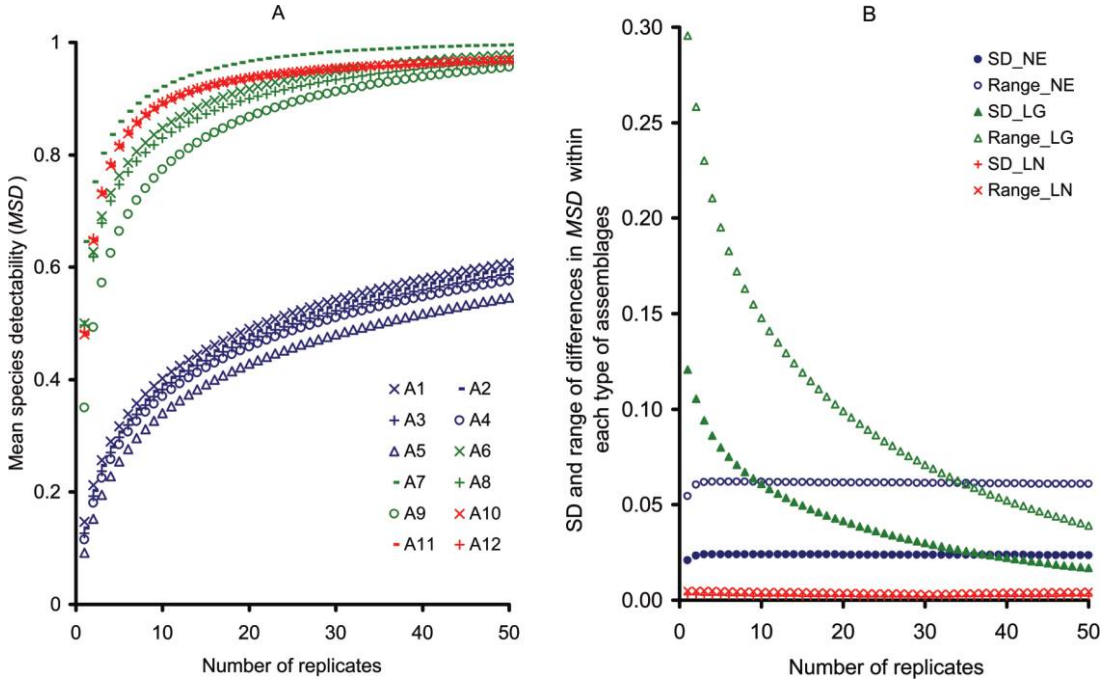

Figure 2: *A*, Mean species detectabilities (MSD) reached by the same number of replicate samples for three types of species-occurrence distributions (SODs): negative exponential (NE; A1–A5), logistic (LG; A6–A9), and linear (LN; A10–A12). *B*, Variability in MSD at a given sampling effort (1 to 50 sample units) for each of the three types of SOD was measured with the range and standard deviation of differences in mean species detectabilities among the assemblages of each type.

assemblages, followed by logistic SOD assemblages and then linear SOD assemblages. The MRS-MSD relationships converged when MRS was equal to or greater than 0.8, which required four to 388 replicates across the 12 assemblages. The average difference in MSD across the three types of SOD was equal to or smaller than 0.24, compared with 0.47 for rarefaction-based estimates and 0.65 for Jack-2 estimates.

The MRS-MSD relationships were also consistent within each type of SOD (fig. 4*A*). For a given MRS value, the difference in MSD was <0.05 for both the negative exponential and linear SOD assemblages and ≤0.07 for logistic SOD assemblages (fig. 4*B*).

Comparing Biases of the Three Methods at the Same Sampling Effort

The MSD based on the seven replicates per assemblage it took to achieve an MRS = 0.6 for the 12 combined assemblages varied from 0.21 to 0.89 across the 12 assemblages (SD = 0.28): 0.47 to 1, based on Jack-2 estimates (SD = 0.24), and 0.34 to 0.59, based on MRS-based standardizations (SD = 0.10; fig. 5). Although MRS standardizations did not remove all of the variability in MSDs among these assemblages, they did result in much less variable MSDs than other standardizations and hence should lead to more accurate estimates of the relative differences in richness among assemblages.

Precision of MRS-MSD Relationships

The MRS-MSD relationships were generally precise except for the negative exponential model at low TSR (A2, TSR = 50, RMSE of MSD ≤0.1; fig. C3). For all other assemblages examined, RMSE ranged between 0.02 and 0.04. The relationships between RMSE and MRS were generally weak and inconsistent across the assemblages examined. The precision of MRS-MSD relationships did not vary significantly with MRS.

Discussion

Data from biological surveys are used to establish spatial patterns or temporal trends in species richness (Rosenzweig 1995) and quantify human impacts on ecosystems (NRC 2000). However, comparisons of species richness among assemblages or locations can be compromised because species detectabilities are not only rarely estimated

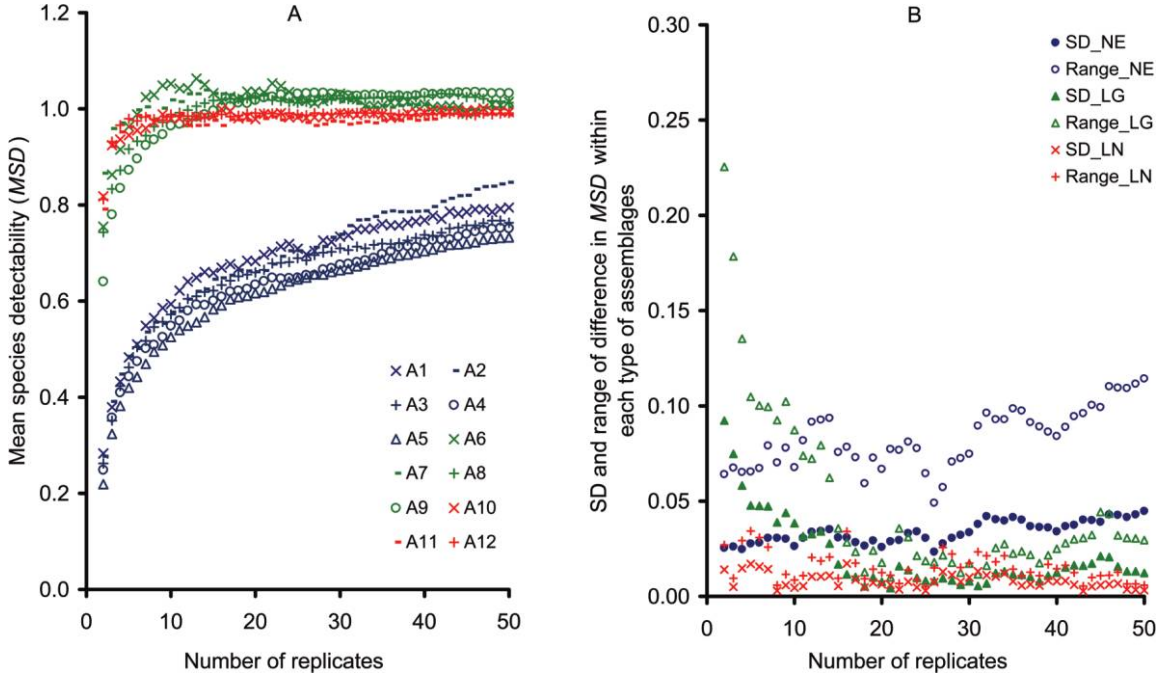

Figure 3: *A*, Mean species detectabilities (MSD) reached by the second-order jackknife estimates of species richness for three types of species-occurrence distributions (SODs): negative exponential (NE; A1–A5), logistic (LG; A6–A9), and linear (LN; A10–A12). *B*, Variability in MSD at a given sampling effort (1 to 50 sample replicates) for each of the three types of SOD was measured with the range and standard deviation of differences in mean species detectabilities among the assemblages of each type.

but also variable across different assemblages. In this study, we showed that three methods of comparing species richness (rarefaction, statistical estimations of TSR, and MRS-based standardization) are differentially sensitive to variation in SODs. The degree to which the type of sample standardization affects inferred patterns of species richness over space or time depends on how diverse compared assemblages are in terms of their species-occurrence distributions. However, unless we can show that the SODs are similar for the assemblages being compared, we should assume they are different and use sample standardizations and estimators of species richness that are least sensitive to differences in SOD. We therefore focus the discussion on (1) how well rarefaction, statistical estimators, and standardization on MRS cope with diverse SODs in comparing species richness and (2) under what conditions different types of SODs are likely to occur when comparing species richness.

Problems with Rarefaction

Rarefaction is widely used to compare species richness among assemblages (Gotelli and Graves 1996; Gotelli and Colwell 2001; Colwell et al. 2004). However, because standardization on a given number of replicates can yield very different MSDs across assemblages (fig. 2), rarefaction must necessarily bias estimates of relative differences among assemblages. This problem was suggested more than 30 years ago when Fager (1972) showed that assemblage evenness, which is influenced by the type of SOD, strongly affected rarefaction curves. However, the severity of this problem for meeting the assumption of equal MSD and thus allowing robust comparisons of species richness has not been adequately appreciated. The rule of thumb that the samples compared should come from taxonomically similar assemblages and similar habitats (Gotelli and Graves 1996) is unlikely to guarantee similar SODs or similar MSDs because taxonomically similar assemblages from similar habitats can vary greatly in their SODs, as we observed in the local fish data (table A2). Disturbances can also significantly change the SODs or species dominance patterns of otherwise similar assemblages (Cao et al. 1998; Mackey and Currie 2001). Because such disturbances can either increase or decrease assemblage evenness, comparisons based on rarefied species richness are unlikely to either accurately or consistently describe the effects of disturbances on species richness (Cao and Hawkins 2005).

Estimates of species richness are also used to rank sites in terms of their species diversity rather than to quantify

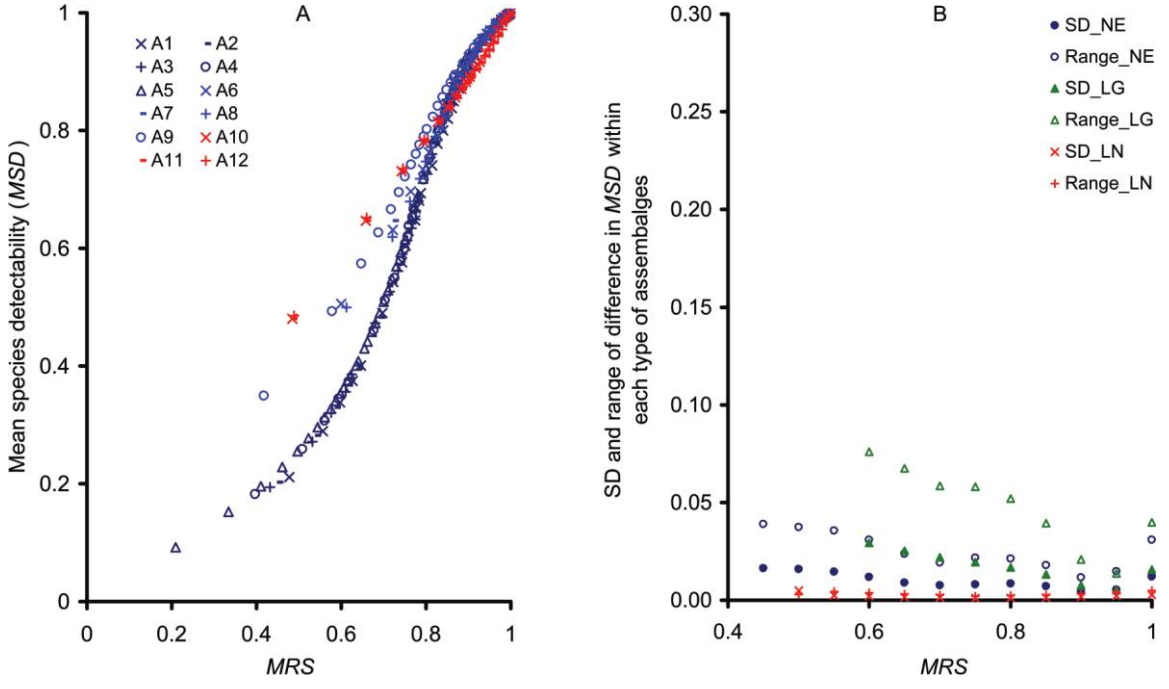

Figure 4: *A*, Relationships between mean species detectability (MSD) indexed as mean replicate similarity (MRS) and the true value (MSD) for the three types of species-occurrence distributions (SODs): negative exponential (NE; A1–A5), logistic (LG; A6–A9), and linear (LN; A10–A12). *B*, Variability in mean species detectability at a given sampling effort (1 to 50 replicates) for each of the three types of SOD was measured with the range and standard deviation of differences in mean species detectabilities among the assemblages of each type.

their relative differences. When different assemblages were associated with a single type of SOD, rarefaction appeared to work well in ranking assemblages based on diversity (fig. C1). However, when assemblages had different types of SODs, ranks can shift with increasing sampling effort, results also observed by Thompson and Withers (2003) in a simulation study. Early reviews of published species accumulation curves derived from field data suggested that curves seldom crossed (Simberloff 1978). However, recent analyses on a variety of assemblages show that species curves do frequently cross one another (e.g., Lande et al. 2000; Scheiner et al. 2000; Cao et al. 2001; Barnosky and Carrasco 2002; Thompson et al. 2003; Chiarucci and Bonini 2005). Other studies that show that the ranks of sites based on raw sample richness can differ from those based on statistical estimates (e.g., Stout and Vandermeer 1975; Hughes et al. 2002) also imply that species-accumulation curves cross at a sampling effort higher than that typically used in field studies.

Rarefaction was developed as a way to consistently compare species richness when sampling effort varies among sampling sites and was based on well-developed statistical models (Hurlbert 1971; Simberloff 1978). However, this method implicitly assumes that a standard sampling effort would reach an equal MSD in different assemblages. This one-size-fits-all assumption is clearly not realistic, and it can yield substantially different MSDs and severely bias inferences regarding relative differences or ranks in species richness (fig. C2).

Limitations in Statistically Estimating True Species Richness

Statistical estimators, such as Jack-2, were designed to estimate TSR when species detectabilities are imperfectly measured (see Colwell and Coddington 1994; Walther and Moore 2005 for reviews). These estimators have been shown to be dependable or useful in some situations (e.g., Colwell and Coddington 1994; Chazdon et al. 1998). Jack-2 estimates produced MSDs that were higher than those obtained from raw samples in this study, which shows that this estimator can account for variable species detectabilities to some extent. However, the Jack-2 estimator not only failed to accurately predict TSR, but it also failed to minimize the difference in MSD among samples standardized on effort from different assemblages (figs. 3*B*, 5). Statistical estimators have been increasingly criticized for poor performance (e.g., Cao et al. 2004; Melo 2004; Green et al. 2005), sometimes severely so. For example, Palmer et al. (2002) asserted that these estimators were "unlikely

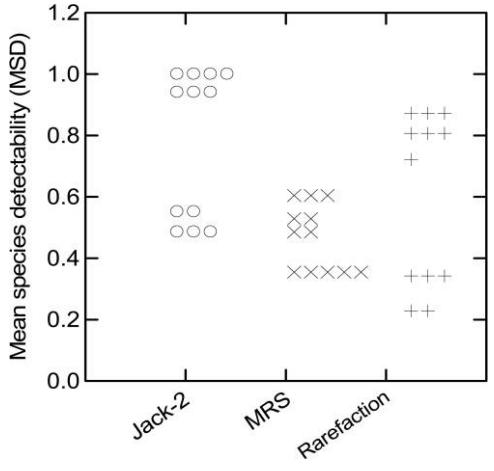

Figure 5: Mean species detectabilities (MSD) achieved by three methods: mean replicate similrity (MRS)-based standardization, the second-order jackknife estimator, and rarefaction for all 12 assemblages based on the same total sampling effort (84 replicates) applied across assemblages, which allowed either seven replicates to be collected for each of 12 assemblages or an MRS value of 0.60 to be reached.

to outperform the guess of experienced botanists" in vegetation surveys, and O'Hara (2005) concluded that their use seemed futile.

Ecologists and statisticians continue to seek more accurate estimators of species richness (e.g., Colwell et al. 2004; Mao and Colwell 2005; Pledger 2005; Dorazio et al. 2006; Hong et al. 2006). Some of the newly developed estimators have been reported to perform less well than certain established ones, for example, Jack-2 (Hortal et al. 2006), and other new estimators have yet to be rigorously tested across a wide range of assemblage data. In general, two fundamental challenges remain in developing robust estimators. First, no single method has been developed that is robust to the effects of SOD or species-abundance distribution on estimates of richness (Colwell and Coddington 1994; Wagner and Wildi 2002; Brose et al. 2003; Dorazio et al. 2006), and none seems likely to emerge (Esty 1986). Second, the practical necessity of taking a relatively small number of samples in field surveys means that estimates of the number of rare species in an assemblage, which represent the majority of species, will be especially problematic. Small samples simply do not contain enough information for a statistical method to accurately estimate the number of rare species present and hence TSR (Good 1953; Mao and Colwell 2005; Curtis et al. 2006; Dorazio et al. 2006). Although we believe it is important to pursue more accurate statistical methods of estimating TSR, we think it may be less problematic to estimate the relative differences in species richness between two or more assemblages.

Circumventing the Variable Detectability Problem by Standardizing on MSD

Standardization on MSD is essential for comparing species richness. However, the effectiveness of this standardization depends on how consistently the specific index of MSD used (MRS in our study) is related to true MSD across different assemblages. In this study, MRS-MSD relationships were highly consistent when assemblages all exhibited any single type of SOD. The relationships were less consistent across assemblages with distinctly different SODs but still much more consistent than for observed or extrapolated species curves (fig. 5) because of the lower variability in MSD across assemblages.

A relatively high MRS, for example, $\geq$0.8, may be required to quantify subtle differences among assemblages in species richness. Such a high MRS value can require extremely high sampling efforts for some assemblages (e.g., 388 replicates for A5) and hence may not be practical when comparisons among multiple assemblages are desired. However, the total sampling effort required in a survey will depend on the range of SODs and TSR involved, and it may be possible to partition sampling in such a way to achieve the required MSD across all assemblages. Because MRS-MSD relationships were generally precise, except for species-poor, negative-exponential-type assemblages (fig. C3), and little affected by the level of MRS, such a standardization should be applicable to real surveys.

The Lincoln-Petersen model we used to estimate MRS or a similar model proposed by Plotkin and Muller-Landau (2002) can be used to calculate the expected value of MRS for a given random sample size only when the exact SOD is known, something unlikely to occur in practice. However, an MRS-based standardization can be performed by applying a sequential sampling scheme (Krebs 1998) to each assemblage; that is, continue sampling until a targeted MRS is reached (Cao et al. 2002*a*, 2002*b*). The exact procedure will depend on the type of assemblage and sampling method used.

Variability of Species-Occurrence Distributions

The general utility of the three standard methods is dependent on how often SODs differ across the assemblages to be compared. We consider three situations.

Comparing local assemblages of the same type. Stream fish assemblages showed diverse SODs (table A2), but it is not clear how commonly SODs differ among the same type of assemblages in general. Research on patch-occupancy distributions in a variety of assemblages (e.g., Gibson et al. 1999) provide insight into this issue. Left-skewed patch-occupancy distributions (i.e., most species are rare) are approximately equivalent to negative exponential

SODs; right-skewed (i.e., most species are common) or bimodal distributions (i.e., most species are either very common or very rare) are equivalent to logistic SODs, and uniform distributions are equivalent to linear SODs. To date, the available evidence suggests that diverse SODs commonly occur in the same type of assemblages (e.g., van Rensburg et al. 2000; Mehranvar and Jackson 2001).

Comparing assemblages of the same type that differ in spatial scale of sampling. Many studies have examined relationships between local and regional species richness in order to infer the relative importance of local and regional processes (e.g., Caley 1997; Strivastava 1999; Witman et al. 2004). If the SODs of assemblages are strongly dependent on the extent of area sampled (e.g., the stream fish assemblages in this study), the observed relationships between local and regional richness are likely to vary unpredictably unless the mean species detectabilities at both scales are comparable.

Comparing assemblages of different types. Ecologists may desire to compare species richness across different taxonomic groups. For example, Heino et al. (2003, 2005) tested taxonomic concordance in diversity patterns among different types of assemblages. However, any real concordance in diversity patterns will be distorted unless MSDs are similar across different types of assemblages being compared. An assumption of similar MSDs in such cases is particularly problematic because SODs appear to differ even more among different taxonomic groups than within a taxonomic group (app. A). Because SODs can differ in any of the three types of species-richness comparisons, an MRS-based standardization may be generally advantageous over either rarefaction or statistical estimators like Jack-2.

In this study, we focused on estimation of the relative differences in species richness among assemblages. However, unequal mean species detectabilities will also certainly bias comparisons of species composition among assemblages, another focus of community ecology (Gauch 1982; Condit et al. 2002; Plotkin and Muller-Landau 2002). The effect of the bias associated with a standard sampling effort on community analysis needs to be examined. The MRS-based standardization should also be compared to statistical extrapolations of assemblage similarity (Chao et al. 2005) regarding their effectiveness in minimizing such bias. Such investigations should lead to better understanding of assemblage patterns.

Conclusions

We systematically examined how well estimates of species richness derived from rarefaction, statistical estimators, and standardization on MRS achieved constant mean species detectabilities in different assemblages, an essential requirement for unbiased estimation of relative differences in species richness among assemblages. We demonstrated that the MSD reached by a standard sample strongly depends on the species-occurrence distribution in an assemblage, which may vary greatly within certain types of assemblages (e.g., local fish assemblages) and often varies across different taxonomic groups. Rarefaction therefore can severely bias comparisons of species richness across assemblages with different types of SODs, a pitfall that has not been widely recognized. Jack-2 and other statistical estimators tested did not appear to be able to minimize differences in MSD among assemblages. Standardization on MRS resulted in variability of MSD that was similar to that produced by rarefaction and Jack-2 when assemblages exhibited a single type of SOD but resulted in much lower variability in MSD when assemblages with different SODs were compared. The MRS-based standardizations therefore yielded the most accurate estimates of relative difference in species richness across all assemblages. The MRS-MSD relationships were somewhat sensitive to the difference in SODs, especially at low-intermediate levels of MRS, although much less strongly so than for sample richness or extrapolated species accumulation curves. A high level of MRS will therefore be needed to quantify subtle differences in species richness among assemblages of diverse SODs.

Acknowledgments

We thank R. K. Colwell, T. A. Crowl, N. J. Gotelli, F. L. He, J. B. Plotkin, D. W. Roberts, and A. L. Sheldon for their constructive comments on early versions of this manuscript. Comments by three anonymous reviewers further improved the manuscript. S. Urquhart suggested combining random uniform distributions with probabilities of occurrence as a way of simulating individual samples. D.P.L. was an employee of the U.S. Environmental Protection Agency (USEPA) during preparation of this manuscript. This research was partially funded by USEPA Science to Achieve Results program grants R-82863701 and R-83059401 and partially by the National Research Council Associateship Program through USEPA Cooperative Agreement CR826388. This document has been subject to USEPA's peer and administrative review, and it has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Literature Cited

Angermeier, P. L., and R. A. Smogor. 1995. Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. Canadian Journal of Fisheries and Aquatic Sciences 52:936–945.

Bailey, L. L., T. R. Simons, and K. H. Pollock. 2004. Estimating site occupancy and species detectability parameters for terrestrial salamanders. Ecological Applications 14:692–702.

Barnosky, A. D., and M. A. Carrasco. 2002. Effects of Oligo-Miocene global climate changes on mammalian species richness in the northwestern quarter of the USA. Evolutionary Ecology Research 4:811–841.

Bell, G. 2001. Neutral macroecology. Science 293:2413–2418.

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.

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.

Bunge, J., and M. Fitzpatrick. 1993. Estimating the number of species: a review. Journal of the American Statistical Association 88:364–373.

Burnham, K. P., and W. S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. Biometrika 65:623–633.

Caley, M. J. 1997. The relationship between local and regional diversity. Ecology 78:70–80.

Cao, Y., and C. P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. Journal of Applied Ecology 42:954–965.

Cao, Y., D. D. Williams, and N. E. Williams. 1998. How important are rare species in community ecology and bioassessment. Limnology and Oceanography 43:1403–1409.

Cao, Y., D. P. Larsen, and R. M. Hughes. 2001. Evaluating sampling sufficiency in fish assemblage surveys: a similarity-based approach. Canadian Journal of Fisheries and Aquatic Sciences 58:1782–1793.

Cao, Y., D. D. Williams, and D. P. Larsen. 2002*a*. Comparisons of ecological communities: the problem of sample representativeness. Ecological Monographs 72:41–56.

Cao, Y., D. P. Larsen, R. M. Hughes, P. Angermeier, and T. Patton. 2002*b*. Sampling efforts affect multivariate comparisons of stream assemblages. Journal of the North American Benthological Society 21:707–714.

Cao, Y., D. P. Larsen, and D. White. 2004. Estimating regional species richness using a limited number of survey units. Ecoscience 11: 23–35.

Carter, J. L., and V. H. Resh. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. Journal of the North American Benthological Society 20:658–682.

Chao, A., R. L. Chadzon, R. K. Colwell, and T.-J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence data and abundance data. Ecology Letters 8:148–159.

Charles, D. F., S. S. Dixit, B. F. Cumming, and J. P. Smol. 1991. Variability in diatom and chrysophyte assemblages and interred pH: paleolimnological studies of Big Moose Lake, New York, USA. Journal of Paleolimnology 5:267–284.

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 northeastern Costa Rica. Pages 285–309 *in* F. Dallmeier and J. A. Comiskey, eds. Forest diversity research, monitoring and modelling. Parthenon, Washington, DC.

Chiarucci, A., and I. Bonini. 2005. Quantitative floristics as a tool for the assessment of plant diversity in Tuscan forests. Forest Ecology and Management 212:160–170.

Chiarucci, A., A. J. Enright, G. L. W. Perry, B. P. Miller, and B. B. Lamont. 2003. Performance of nonparametric species richness estimators in a high diversity plant community. Diversity and Distribution 9:283–295.

Colwell, R. K. 2005. EstimateS: statistical estimation of species richness and shared species from samples. Version 7.5. User's guide and application published at http://viceroy.eeb.uconn.edu/EstimateS.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of Royal Society B: Biological Sciences 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.

Condit, R., S. P. Hubbell, J. V. LaFrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. Ashton. 1996. Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. Journal of Ecology 84:549–562.

Condit, R., N. Pitman, E. G. Leigh, J. Chave, J. Terborgh, R. A. Foster, P. Nunez, et al. 2002. Beta-diversity in tropical forest trees. Science 295:666–668.

Cormack, R. M. 1989. Log-linear models for capture-recapture. Biometrics 45:395–413.

Curtis, T. P., I. M. Head, M. Lunn, S. Woodcock, P. D. Schloss, and W. T. Sloan. 2006. What is the extent of prokaryotic diversity? Philosophical Transactions of the Royal Society B: Biological Sciences 361:2023–2037.

Dorazio, R. M., J. A. Royle, B. Soderstrom, and A. Glimskar. 2006. Estimating species richness and accumulation by modeling species occurrences and detectability. Ecology 87:842–854.

Esty, W. W. 1986. Estimation of the size of a coinage: a survey and comparison of methods. Numismatic Chronicle 146:185–215.

Fager, E. W. 1972. Diversity: a sampling study. American Naturalist 106:293–310.

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.

Gaston, K. J. 1996. Species richness: measure and measurement. Pages 76–113 *in* K. J. Gaston, ed. Biodiversity: a biology of number and difference. Blackwell Science, Oxford.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Gibson, D. J., J. S. Ely, and S. L. Collins. 1999. The core-satellite species hypothesis provides a theoretical basis for Grime's classification of dominant, subordinate, and transient species. Journal of Ecology 87:1064–1067.

Good, L. J. 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40:237–264.

Goodall, D. W. 1978. Numerical methods of classification. Pages 250–287 *in* R. H. Whittaker, ed. Classification of plant communities. Junk, The Hague.

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity:

procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391.

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

Gray, J. S. 1978. The structure of meiofauna communities. Sarsia 64: 265–272.

Green, J. L., A. Hastings, P. Arzberger, F. J. Ayala, K. L. Cottingham, K. Cuddington, F. Davis, et al. 2005. Complexity in ecology and conservation: mathematical, statistical, and computational challenges. BioScience 55:501–510.

Hanski, I., and M. Gyllenberg. 1993. Two general metapopulation models and core-satellite hypothesis. American Naturalist 142:17–41.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10: 1456–1477.

He, F. L., and P. Legendre. 2002. Species diversity patterns derived from species-area models. Ecology 83:1185–1198.

He, F. L., P. Legendre, and J. V. MacFrankie. 1997. Distribution patterns of tree species in a Malaysian tropical forest. Journal of Vegetation Science 8:105–114.

Heino J., T. Muotka, R. Paavola, and L. Paasivirta. 2003. Among-taxon congruence in biodiversity patterns: can stream insect diversity be predicted using single taxonomic groups? Canadian Journal of Fisheries and Aquatic Sciences 60:1039–1049.

Heino J., R. Paavola, R. Virtanen, and T. Muotka. 2005. Searching for biodiversity indicators in running waters: do bryophytes, macroinvertebrates, and fish show congruent diversity patterns? Biodiversity and Conservation 14:415–428.

Hellmann, J. J., and G. W. Fowler. 1999. Bias, precision, and accuracy of four measures of species richness. Ecological Applications 9: 824–834.

Hong, S.-H., J. Bunge, S.-O. Jeon, and S. S. Epstein. 2006. Predicting microbial species richness. Proceedings of the National Academy of Sciences of the USA 103:117–122.

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.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, S. S. Intelmann, S. C. Corbett, M. C. Arbogast, and R. C. Hjort. 2002. Electrofishing distance needed to estimate species richness in raftable Oregon rivers. North American Journal of Fisheries Management 22:1229–1240.

Hurlbert, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586.

Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579.

Kery, M., and H. Schmid. 2006. Estimating species richness: calibrating a large avian monitoring programme. Journal of Applied Ecology 43:101–110.

Krebs, C. J. 1998. Ecological methodologies. 2nd ed. Cummings, New York.

Lande, R. P., J. DeVris, and T. R. Walla. 2000. When species accumulation curves intersect: implications for ranking diversity using small samples. Oikos 89:601–605.

Legendre, P., and Legendre, L. 1998. Numerical ecology. 2nd ed. Elsevier, New York.

Lenat, D. R. 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water-quality rating. Journal of North American Benthological Society 12:279–290.

MacArthur, R. H. 1965. Patterns of species diversity. Biological Review 40:510–533.

MacKenzie, D. I., and W. L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? Ecology 83:2387–2393 (erratum, 83:3532).

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modelling: inferring patterns and dynamics of species occurrence. Elsevier, New York.

Mackey, R. L., and D. J. Currie. 2001. The diversity-disturbance relationship: is it generally strong and peaked? Ecology 82:3479–3492.

Mao, C. X., and R. K. Colwell. 2005. Estimation of species richness: mixture models, the role of rare species, and inferential challenges. Ecology 86:1143–1153.

May, R. M. 1975. Patterns in multi-species communities. Pages 142–162 *in* R. M. May, ed. Theoretical ecology: principles and applications. Saunders, Toronto.

McGeoch, M. A., and K. J. Gaston. 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. Biological Reviews 77:311–331.

Mehranvar, L., and D. A. Jackson. 2001. History and taxonomy: their roles in the core-satellite hypothesis. Oecologia (Berlin) 127:131–142.

Melo, A. S. 2004. A critique of the use of jackknife and related nonparametric techniques to estimate species richness. Community Ecology 5:149–157.

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.

O'Hara, R. B. 2005. Species richness estimators: how many species can dance on the head of a pin? Journal of Animal Ecology 74: 375–386.

Palmer, M. W., P. G. Earls, B. W. Hoagland, P. S. White, and T. Wohlgemuth. 2002. Quantitative tools for perfecting species lists. Environmetrics 13:121–137.

Patton, T. M., W. A. Hubert, F. J. Rahel, and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. North American Journal of Fisheries and Management 20:394–398.

Pledger, S. 2005. The performance of mixture models in heterogeneous closed population capture-recapture. Biometrics 61:868–876.

Plotkin, J. B., and H. C. Muller-Landau. 2002. Sampling the species composition of a landscape. Ecology 83:3344–3356.

Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13: 105–119.

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology 29:254–282.

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

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

Scheiner, S. M., S. B. Cox, M. Willig, G. G. Mittelbach, C. Osenberg, and M. Kaspari. 2000. Species richness, species-area curves and Simpson's paradox. Evolutionary Ecology Research 2:791–802.

Simberloff, D. 1978. Use of rarefaction and related methods in ecology. Pages 150–165 *in* K. L. Dickson, J. Cairns Jr., and R. J. Livingston, eds. Biological data in water pollution assessment: quantitative and statistical analyses. American Society for Testing and Materials, Philadelphia.

Stout, J., and J. Vandermeer. 1975. Comparisons of species richness for stream-habitating insects in tropic and mid-latitude streams. American Naturalist 109:263–280.

Strivastava, D. S. 1999. Using local-regional richness plots to test for species saturation: pitfalls and potentials. Journal of Animal Ecology 68:1–16.

Thompson, G. G., and P. C. Withers. 2003. Effect of species richness and relative abundance on the shape of the species accumulation curve. Austral Ecology 28:355–360.

Thompson, G. G., P. C. Withers, E. R. Pianka, and S. A. Thompson. 2003. Assessing biodiversity with species accumulation curves: inventories of small reptiles by pit-trapping in Western Australia. Austral Ecology 28:361–383.

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.

Ulrich, W., and M. Ollik. 2005. Limits to the estimation of species richness: the use of relative abundance distributions. Diversity and Distributions 11:265–273.

van Rensburg, B. J., M. A. McGeoch, W. Matthews, S. L. Chown, and A. S. van Jaarsveld. 2000. Testing generalities in the shape of patch occupancy frequency distributions. Ecology 81:3163–3177.

Wagner, H. H., and O. Wildi. 2002. Realistic simulation of the effects of abundance distribution and spatial heterogeneity on non-parametric estimators of species richness. Ecoscience 9:241–250.

Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision, accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 28:815–829.

Watson, D. M. 2003. The "standardized search": an improved way to conduct bird surveys. Austral Ecology 28:515–525.

Whittaker, R. H. 1975. Communities and ecosystems. 2nd ed. Macmillan, New York.

Williams, B. K., J. D. Nicholls, and M. J. Conroy. 2001. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, New York.

Williams, C. B. 1964. Patterns in the balance of nature and related problems in quantitative ecology. Academic Press, London.

Winkler, H., and C. Kampichler. 2000. Local and regional species richness in communities of surface-dwelling grassland Collembola: indication of species saturation. Ecography 23:385–392.

Witman, J. D., R. J. Etter, and F. Smith. 2004. The relationship between regional and local species diversity in marine benthic communities: a global perspective. Proceedings of the National Academy of Sciences of the USA 101:15664–15669.

Wolda, H. 1981. Similarity indices, sample size and diversity. Oecologia (Berlin) 50:296–302.

Wright, J. F., R. J. Blackburn, M. Gunn, M. T. Furse, and P. D. Armitage. 1996. Macroinvertebrate frequency data for the RIVPACS III sites in Great Britain and their use in conservation evaluation. Aquatic Conservation: Marine and Freshwater Ecosystems 6:141–167.

Associate Editor: Steven L. Chown
Editor: Donald L. DeAngelis