Estimating Diversity in Unsampled Habitats of a Biogeographical Province

MICHAEL L. ROSENZWEIG,*‡ WILL R. TURNER,* JONATHAN G. COX,*
AND TAYLOR H. RICKETTS†

*Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, AZ 85721–0088, U.S.A.
†Department of Biological Sciences, Stanford University, Stanford, CA 94305–5020, U.S.A.

Abstract: *Estimating the number of species in a biogeographical province can be problematic. A number of methods have been developed to overcome sample-size limits within a single habitat. We evaluated six of these methods to see whether they could also compensate for incomplete habitat samples. We applied them to the butterfly species of the 110 ecoregions of Canada and the United States. Two of the methods use the frequency of species that occur in a few of the sampled ecoregions. These two methods did not work. The other four methods estimate the asymptote of the species-accumulation curve (the graph of "number of species in a set of samples" versus "number of species occurrences in those samples"). The asymptote of this curve is the actual number of species in the system. Three of these extrapolation estimators produced good estimates of total diversity even when limited to 10% of the ecoregions. Good estimates depend on sampling ecoregions that are hyperdispersed in space. Clustered sampling designs ruin the usefulness of the three successful methods. To ascertain their generality, our results must be duplicated at other scales and for other taxa and in other provinces.*

Key Words: biodiversity, butterfly, ecoregion, habitat heterogeneity, species diversity

Estimación de la Diversidad en Hábitats no Muestreados de una Provincia Biogeográfica

Resumen: *La estimación del número de especies en una provincia biogeográfica puede ser problemático. Se ha desarrollado un número de métodos para superar los límites del tamaño de muestra dentro de un solo hábitat. Evaluamos seis de estos métodos para ver si podrían compensar por muestras incompletas de hábitat. Aplicamos estos métodos a especies de mariposas de las 110 ecoregiones de Canadá y los Estados Unidos. Dos de los métodos usan la frecuencia de las especies que ocurren en algunas de las ecoregiones muestreadas. Estos dos métodos no sirvieron. Los otros cuatro métodos estimaron la asíntota de la curva de acumulación de especies (la gráfica de "el número de especies en un juego de muestras" contra el "número de ocurrencias de especies en éstas muestras"). La asíntota de ésta curva es el número real de especies en el sistema. Tres de éstos estimadores de extrapolación produjeron buenas estimaciones de la diversidad total aún cuando se limitaron al 10% de las ecoregiones. Las buenas estimaciones dependen del muestreo de ecoregiones altamente dispersas en el espacio. Los diseños de muestreos en agrupamientos arruinan la utilidad de los tres métodos exitosos. Para asegurar su generalidad, nuestros resultados deben ser duplicados a otras escalas y para otros taxones en otras provincias.*

Introduction

Counting the number of species, *S*, in a heterogeneous region presents two distinct sampling problems: every real sample includes (1) only a finite number of individuals and (2) only a finite number of habitats. Hence, enumerations of species fall short both because we have not counted every individual and because we have not looked in every place.

The first problem is the classic sample-size problem (Fisher et al. 1943). Because real-life samples are limited, the next individual we collect from a place could

‡*email scarab@u.arizona.edu*
Paper submitted June 18, 2001; revised manuscript accepted September 24, 2002.

come from a species we had not before seen there. Thus, the number of species actually collected is most often smaller than the number actually present. It cannot be greater, so the raw number we sample is a negatively biased estimate of the true number present. Methods for dealing with this problem have achieved considerable sophistication and success.

The second problem arises from the fact that every finite set of samples is only a subset of available habitats. Because each species also lives in a restricted set of habitats, the only way we can be sure of having a chance to detect all species is to sample everywhere. In practice, that is not possible.

We dealt with the second problem and hypothesized that the methods advanced for dealing with the sample-size problem may also be successful in dealing with the heterogeneity problem. We evaluated such methods based on their ability to perform accurately and reliably with as small a sample set as possible. (Small samples require the least amount of money and time to obtain and are often the only ones available.) Our results suggest that Holdridge et al.'s (1971) method and two other extrapolation formulas based on this method—but quite different in detail—may be able to compensate for incomplete habitat sampling.

We dealt only with estimating the number of species, not their relative abundances. We use the term *species diversity* or simply *diversity* to mean the number of species. The "number of kinds" is diversity's original meaning, and we believe it should be restored. Many authors today use *diversity* to mean one of a variety of combinations of the number of species with "evenness." Evenness is a property of the abundance distribution (Hill 1973). Such combinations have led to no useful advances of which we are aware. Besides, evenness deserves to be and can be studied by itself (Smith & Wilson 1996). Finally, the term *species richness* creates another bit of jargon that does nothing to aid our communication with the dedicated laypeople who care about diversity.

Techniques for Dealing with Sample-Size Bias

Fisher himself suggested the first technique for addressing the sample-size problem. He derived an index of diversity independent of sample size called Fisher's alpha (Fisher et al. 1943). But precisely because it is an index, Fisher's alpha side-steps the problem of estimating the number of species itself.

Others have faced the diversity problem squarely. Two of their methods are particularly promising, and we tested their usefulness for dealing with the problem of heterogeneity. Burnham and Overton (1979) derived a distribution-free, jackknife method for estimating S. Lee and Chao (1994) invented another, termed the incidence-based coverage estimator (ICE) to do the same. These two methods remove most of the bias caused by finite sample size (e.g., Colwell & Coddington 1994; Chazdon et al. 1998; Poulin 1998; Hellmann & Fowler 1999). (Note: Most of those who use the jackknife estimator use only its second order, but we use all five in the manner originally prescribed by Burnham and Overton [1979].)

Among the properties shared by ICE and the jackknife method is one that sets them apart from another class of diversity estimators. Although both involve accumulating species by sampling many locations, their actual estimates come from looking at the accumulated total diversity and the number of locations in which each species occurs, not from an examination of the regularities of the accumulation.

In contrast, Holdridge et al. (1971) invented the strategy of extrapolating diversity to its asymptote. The asymptote is the number of species in an infinitely large sample (Palmer 1990; Soberón & Llorente 1993). As more individuals are sampled, diversity accumulates following a relatively smooth, convex, upward line. Holdridge reasoned that the shape of the line contains the information required to specify its asymptote.

To use Holdridge's asymptotic method, one needs a functional form. This must be a skeletal equation capable of fitting many data sets given an appropriate choice of coefficients. Holdridge chose the Michaelis-Menten formula (known among students of predation as the Holling type-II functional response). Used for the purpose of extrapolating a diversity estimate, the formula appears thus:

$$S_{\text{obs}} = S\frac{N}{N+a}, \tag{1}$$

where N is the number of individuals in the sample; a, the half-saturation coefficient, is a coefficient of curvature; S is the asymptote (i.e., the true number of species in the system); and S_{obs} is the number of species in the sample. We abbreviate Holdridge's Michaelis-Menten method as MM. (Fig. 1 shows an example.)

We used Eq. 1 directly, fitting the simulation runs with a nonlinear regression algorithm. To execute MM, the preference of researchers of diversity estimation has been the Eadie-Hofstee formula instead of curve fitting (Colwell & Coddington 1994). Our software (Turner et al. 2000) calculates the estimators with both methods, but the Eadie-Hofstee formula can behave poorly. When it makes estimates using small amounts of data, it exhibits a substantial negative bias. It often actually returns large negative estimates of diversity. And when it provides estimates based on large amounts of data, it often "predicts" the existence of fewer species than are present in the sample.

A New Family of Extrapolation Formulas

Like Eq. 1, an appropriate extrapolation formula rises with a declining slope to an asymptote equal to actual di-

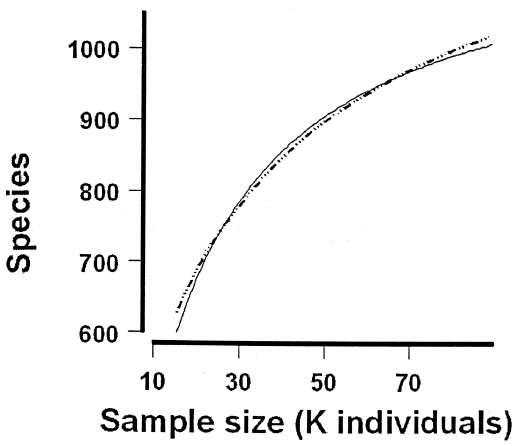

*Figure 1. An example of the use of the Michaelis-Menten formula {*y = P(x/(x+a))*} to estimate diversity, where* P *is the asymptote (and estimate of diversity) and* a *is the curvature. We loaded a total of 89,596 insects collected from Minnesota habitats into a computer database. The set had 1167 species. We sampled repeatedly from the entire set (with replacement) so as to fashion an accumulation curve. The values plotted (broken line) are averages over all 50 runs. Although the total set contained 1167 species, the average sample of 89,596 individuals accumulated only 1018 of them. The Michaelis-Menten curve (thin solid line) fit the sample curve well (*$R^2 = 0.994$*), but showed systematic deviation. The* P *was 1049 species, which is 118 species fewer than were actually present. Data were supplied by E. Siemann (Siemann et al. 1996).*

versity. It must also begin at the point (1,1), however, because when our sample contains only one individual the formula should tell us that it contains one species. This is also true if the only thing we know about is the presence of one species. Equation 1 does not satisfy this criterion because instead of going through the point (1,1) it goes through the point (1,{$S/1+a$}).

A family of formulas that do go through the point (1,1), that rise with a declining slope, and that converge on a positive asymptote is

$$S_{\text{obs}} = S^{1-N^{-f(N)}}, \qquad (2)$$

where $f(N)$ is any positive, unbounded, monotonically increasing function of N. As N rises toward infinity, Eq. 2 converges on S; that is, the asymptote of Eq. 2 is S, the true diversity of the system.

We have worked with many such functions $f(N)$. Pilot results led us to pursue three of them in this study. We substituted them into Eq. 2 to produce three extrapolation estimators: F3, F5, and F6 (names follow those in our lab notes): F3 uses

$$f(N) = q \ln N,$$

F5 uses

$$f(N) = qN^q,$$

and F6 uses

$$N^{-f(N)} = a^{1-N^q}.$$

General Method

We used a real data set: the 561 butterfly species of Canada and the United States (excluding Hawaii and Puerto Rico). This region consists of 110 separate ecoregions. We have the list of species for each of the ecoregions (Ricketts et al. 1999). The popularity and showiness of the taxon means these lists are reasonably complete. Thus, we avoided the first sampling problem: incomplete knowledge of which species live in the sampled places.

We organized the data into a matrix with 561 columns (species) and 110 rows (ecoregions). Each entry in the matrix is a 0 if that species is absent from that ecoregion or a 1 if it is present. Our software (Turner et al. 2000) is designed to sample such matrices, adding ecoregions one by one and analyzing the partial information at each step.

The software sees an occurrence of a species in an ecoregion as one individual. It uses these quasi-individuals to determine such things as the number of singletons (species found in only one ecoregion of the set), double tons (species found in two), and so on for those estimators that require these statistics. It also uses the number of occurrences as sample size (i.e., N) in making use of MM, F3, F5, and F6.

As each ecoregion was added to the sample, the program determined the estimated total diversity produced by each of the six methods listed above. Because there are 561 butterfly species, the correct answer was always 561. Thus, we could determine the success of each method at each step. (Software settings are available at www.evolutionary-ecology.com/data/butterfly.pdf.)

Sampling Strategies

Random Strategies

The computer selects a set of r ecoregions at random. In this random set, some species may be represented more than once and others not at all. Here we repeated each selection of r ecoregions 50 times to obtain average results and their dispersions.

Nonrandom Strategies

Our results showed that the random strategy worked extremely well. One could claim that the random strategy cheats a bit, however, because at each step it selects its r ecoregions from all 110 ecoregions. Hence, no matter how few ecoregions are in a subset r of the 110, the random method is not limited to the subset but is looking at the entire set of 110 ecoregions. On the other hand, each estimate comes only from the data of the subset of r ecoregions. How can selecting from all 110 cause a problem?

The answer is that because all 110 ecoregions are available each time a subset is chosen, a substantial number of replications will reveal with exaggerated accuracy the average number of species in a subset of r ecoregions. So the random strategy ought to generate low variances, smooth accumulation curves, and estimates of total diversity that are perhaps better than can be realistically expected from real-world data sets (where, at each step, only the first r ecoregions would actually be known).

To see that, imagine an explorer who knew nothing of any ecoregion except those he had visited. Especially when he had visited only a few, he might well have accumulated an atypical number of species. A small number would lead to low estimates of the asymptote; a large number would lead to high estimates. Therefore, before we can gain any confidence in the method, we have to limit the computer to the sort of knowledge the explorer would have. We devised several nonrandom strategies to mimic such an explorer and his increasing knowledge.

We began by specifying the geographical position of each ecoregion in a square, virtual space with x and y coordinates ranging from 1 to 110. We assigned each ecoregion a unique latitudinal integer on the y interval. The ecoregion located farthest north received the 1. The ecoregion whose northern-most point was farthest south received the 110. Proceeding from west to east, we similarly assigned longitudinal integers to each ecoregion (Fig. 2).

Next we chose ecoregions from which our fictional explorer would begin surveying. We selected 21 different starting points, which we called *kernels*: four each from among the northern, southern, eastern, and western ecoregions and five from ecoregions in the middle of the continent. Thus, we imagined 21 different explorers and tried to find out how well each was likely to do. Then we estimated the uncertainty they had to be prepared to accept in return for not surveying the entire continent. (Details of how we chose the kernels are available at www.evolutionary-ecology.com/data/butterfly.pdf.)

We next produced ordered lists of the 110 ecoregions for the explorer to follow. Each began at one of the kernels. At each step, our estimators were constrained to proceed strictly in order through a single list. Thus, they could not take advantage of any information except that available from the limited subset of ecoregions.

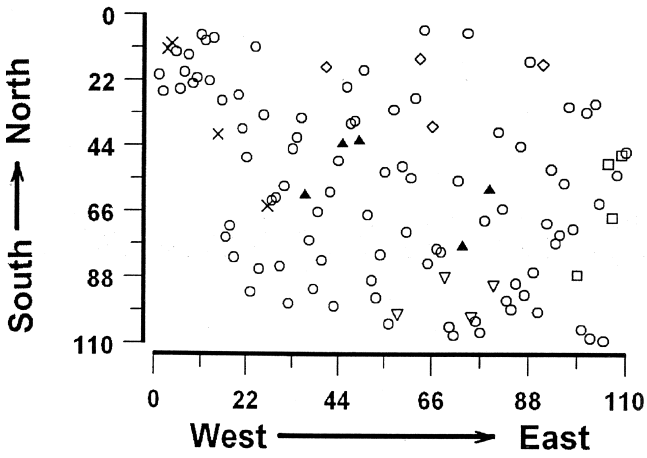

Figure 2. Position of the ecoregions (0–110) in the virtual space (see text). Circles mark ecoregions not selected as kernels. Symbols for ecoregions designated as kernels: boxes, eastern; diamonds, northern; X, western; downward-pointing triangles, southern; solid upward-pointing triangles, central. The entire set roughly resembles the combined shapes of Canada and the continental United States.

We produced three separate ordered lists for each kernel. Each corresponded to one of the following three exploration tactics: blob, cluster, and spread. For the blob tactic we imagine our explorer extending her knowledge outward gradually and without skipping over ecoregions. One may liken this tactic to a drop of liquid permeating a blotter in all possible directions. To execute the blob tactic, we calculated the distances from a kernel to the other 109 ecoregions. Then we ordered the distances from smallest to largest. The ordered list of ecoregions associated with these distances became the blob list for that kernel.

For the cluster tactic we imagine our explorer leaving colonies of explorers behind wherever she goes. At each step, she surveys whichever ecoregion is closest to one of these colonies. Execution of the cluster tactic requires the same set of distances calculated for the blob tactic. At each step of the cluster tactic, however, we calculate the distances from all the chosen ecoregions to all the unchosen ones and then pick the shortest distance. The cluster tactic is similar to the blob tactic because at each step no gaps are allowed between chosen ecoregions. But in practice the cluster tactic penetrated the continent more rapidly than the blob tactic.

For the spread tactic we imagine that at each step our explorer deliberately surveys the largest unexplored part of the continent. Hence, at each step the ecore-

gions that have already been surveyed have intentionally been spread out rather uniformly throughout the continent. In probabilistic terms, they are hyperdispersed, the distances separating them from their nearest neighbors being less variable than they would be if they were chosen at random. We achieved hyperdispersion as follows. The second ecoregion in each spread list was the one farthest from its kernel ecoregion. To order the remaining 108 ecoregions, we projected the east-west positions of the already chosen regions onto the x-axis. Then we found the largest x gap between selected ecoregions and calculated the x value in the center of it. The ecoregion with that value became number three. Next we found the largest gap on the y-axis between selected ecoregions and calculated the y value in the center of it. That ecoregion became number four. We repeated the latter two steps until all 110 ecoregions had been placed in the sampling order. We do not claim that this algorithm achieves maximal hyperdispersion at any stage of ordering the list, but it does deviate from randomness in the direction of substantial hyperdispersion.

Results

Random Accumulation

The number of species observed (S_{obs}) rose gradually toward 561 as more ecoregions were included in the sample. Figure 3 shows the results for one run of 50 replications beginning with a random seed of three, but all runs resembled this one. The jackknife and the ICE—the two estimators that cannot take advantage of the shape of this curve—tracked S_{obs} fairly closely. Their estimates of how many species reside in the entire system were only a bit higher than the number residing in the already surveyed ecoregions. This was not true of the four extrapolation estimators.

F6 was the least successful extrapolation estimator. It was the most erratic, especially when permitted to use fewer than about one-third of the ecoregions. With one-third or more, it flattened out to a modest overestimate of the number of butterfly species. It estimated 590 species when given the data of all 110 ecoregions to work with.

We have confidence that this overestimate is an inherent property of F6. We are less sure whether its erratic behavior on the left side of Fig. 3 is an inherent property of the formula or derives from the software. The software performs nonlinear regression automatically in order to fit F3, F5, F6, and MM. Because F6 has two parameters of curvature, it is the most delicate. Inappropriate automatic choice of initial parameter values can thwart the attempt of the program to converge to a satisfactory nonlinear regression. Checking this possibility requires extensive manual analyses whose results would be of little value.

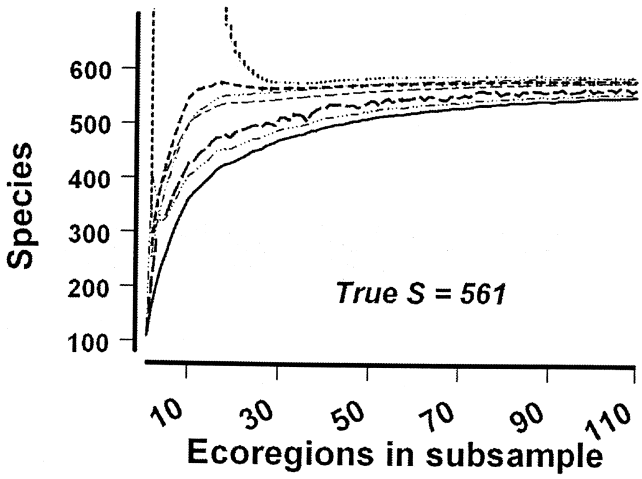

Figure 3. Average results of species accumulation and six diversity estimators (50 runs with random shuffling of ecoregion order and of occurrences within each successive set of chosen ecoregions). Result type is most readily distinguished over x *= 15, where the following* y *values were obtained: accumulation = 369 species; incidence-based coverage estimator (ICE) = 410 species; jackknife method = 431; Michaelis-Menten (MM) method = 508; F3 = 512; F5 = 556; F6 off-scale at 1456 species. All estimators converge over large sample sizes. Unlike the MM curve in Fig. 1, the extrapolation estimators do not follow the numbers of species in the accumulation curve. Instead, the software utilizes that curve up to a certain number (*r*) of ecoregions, performs a fit like that of Fig. 1, then reports the results of its extrapolation as the* y *value over that* r *value. Next, it adds another ecoregion, refits the data, and reports the results of that extrapolation. Results changed little when the random seed was altered or the number of runs was increased greatly. Table 1 reports statistics for a slightly different protocol with similar results.*

MM and F3 did much better than F6. On the left side of Fig. 3, their results fluctuated negligibly, especially compared with those of F6 (Table 1). With moderate numbers (11–50) of ecoregions in the pooled data, MM showed slightly more negative bias than F3. Using all 110 ecoregions, MM estimated 579 butterfly species, the most accurate of any estimator. With all 110 ecoregions, F3's estimate was 588 species. But the end estimate is not very useful. With all 110 ecoregions in hand, one already knows about virtually all the species and no longer needs an estimator.

Where information is least complete and a successful estimator would be most valuable, F5 did best. F5 reached 546 species with the use of only 10 ecoregions. Its terminal estimate was 583 species, so it was also the flattest estimator—the one least affected by the proportion of ecoregion lists included. With the use of 11 ecoregions,

Table 1. Extrapolation estimates of species diversity with 11 or 22 ecoregions chosen randomly from the full set of 110.*

	Includes first 11 ecoregions				*Includes first 22 ecoregions*			
	F6	*F5*	*F3*	*MM*	*F6*	*F5*	*F3*	*MM*
Minimum	328.4	236.8	234.2	244.7	354.9	418.7	412.	413.0
Maximum	3885.6	2660.2	1050.3	915.6	1628.2	840.9	730.	674.7
Mean	905.0	668.9	535.9	524.5	749.3	583.4	560.6	542.7
Median	622.0	545.4	512.4	514.1	650.6	566.5	557.5	542.8
SD	762.6	480.6	165.7	133.6	322.9	94.3	76.2	17.8

**We ran 50 replicates, saved each one, and then averaged the 50 estimates. Results pictured in Fig. 3 come from a slightly different method. A perfect estimate would have been 561, the total number of species in the set of ecoregions. The median of F5 was the best estimate of diversity.*

the estimates were F5 = 556, F3 = 512, and MM = 508 (Fig. 3).

In Fig. 3 and similar runs, we asked the software to have the estimators fit the average values of observed *S*. This gave a good idea of the patterns exhibited by the mean extrapolation estimates. But it did not report the variation in those estimates for any single *r* value (*r* = number of ecoregions included in the estimate). To examine such variation, we ran 50 replicates and saved each one separately (Table 1). The means of these results were somewhat different from those of Fig. 3 because the latter uses means for the nonlinear regressions, whereas the former regresses the separate results and then calculates the mean estimate afterward.

At both *r* = 11 and *r* = 22, MM produced the least variability, F3 somewhat more, and F5 more still. The variability of F6 was large even at *r* = 22.

The means and medians produced by separating the regressions ranked similarly to those obtained from the previous single-regression method (Table 1). The median of F5 at *r* = 11 remained superior to the others, but its mean seemed to have acquired a substantial positive bias. At *r* = 22, the median of F3 was slightly closer to 561 than that of F5, but probably not significantly so. In any case, F5's estimates of the mean using the single-regression technique were the best of all. Thus, there would seem to be no need in practice to go to the considerable extra trouble of performing separate regressions.

Nonrandom Accumulation

As in the case of random accumulation, both the ICE and the jackknife returned estimates only modestly higher

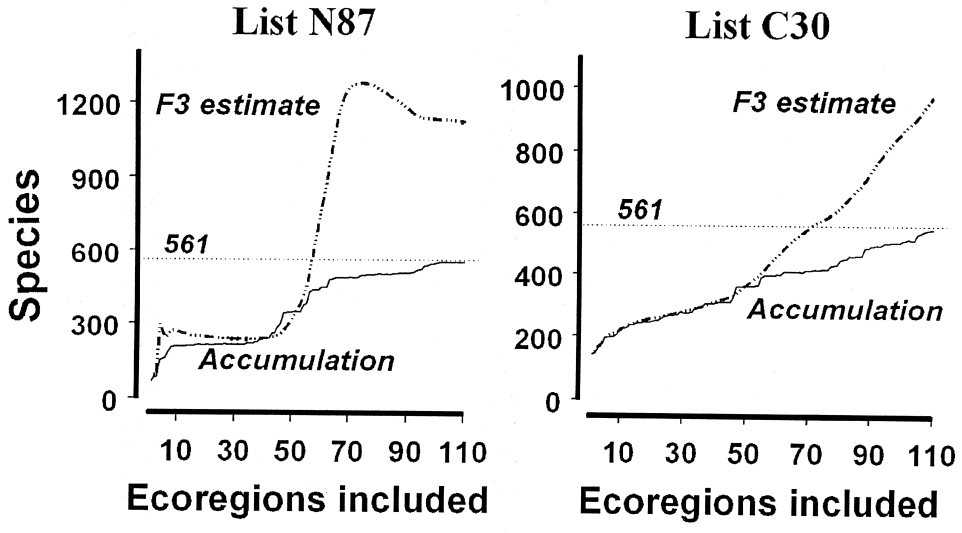

Figure 4. Two examples of F3 analyses using one cluster list (N87) and one blob list (C30). These sorts of lists (see text) accumulate species irregularly. The N87 cluster list accumulated species rapidly in the vicinity of its kernel and appeared to be approaching an asymptote. Then, at about x *= 43, it tapped a group of ecoregions with a novel set of species. It rose abruptly and the estimator, F3, responded with large overestimates. The C30 blob list accumulated species more steadily and never seemed to approach an asymptote. That led to steadily rising F3 overestimates. The F5 and Michaelis-Menten analyses of N87 and C30 behaved similarly.*

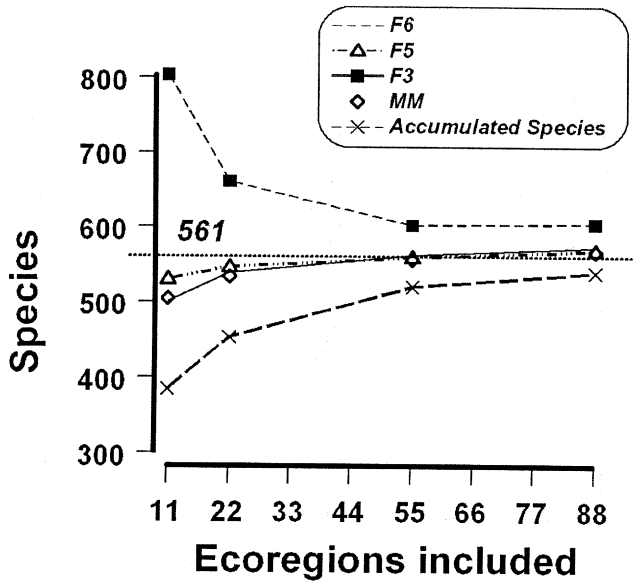

Figure 5. Trends in the means of the extrapolation estimators. The data come from analyses of the 21 spread lists, of which Fig. 6 gives one example. Michaelis-Menten (MM), F5, and F3 all perform well. F5 is the flattest and least biased overall. MM and F3 are so close that we plotted only the line for F3 and the symbols for MM.

than the actual number of species observed at each step. They did this no matter how irregular the shape of the raw accumulation curve because neither of these estimators depends on the shape of that curve. The other four estimators do, however, and their success depends on the accumulation curve rising reliably toward the true asymptote.

When ecoregions were sampled in fixed order instead of random order, the accumulation curve of observed species did not always resemble a smooth curve approaching an asymptote. In particular, both cluster and blob tactics produced shapes quite different from a smooth approach to an asymptote (Fig. 4). They did so because they focus their early samples in a single portion of the continent.

For some of these lists, the curve of observed species rapidly accumulated the species of that region and leveled off because the ecoregions being added were nearby and rather similar in species composition. That circumstance created an early, false appearance of leveling off. The extrapolation estimators detected it and, at first, returned very low estimates of total diversity. Beyond some r value, the ecoregion lists being added began to tap into ecoregions whose species overlapped very little with those previously chosen. Observed diversity began to climb more rapidly. The raw curve's accumulation rate got steeper, and the extrapolation estimators revised their estimates upward.

For other lists, gradual expansion outward from a kernel ecoregion led to steady, nearly linear growth in the number of species observed (Fig. 4). This too caused all extrapolation estimators to behave badly. They were looking for an asymptote that always seemed farther and farther away.

In short, no estimator did a useful job when it was based on blob or cluster tactics of ecoregion exploration.

The spread tactic did very well, however. Accumulation curves based on these lists were not so irregular. Hence, as with random accumulation, F3, F5, and MM all succeeded. On average, they had already achieved a most reasonable average estimate of total diversity with information from only about 10% (r = 11) of the ecoregions in a list (Fig. 5). Spread list W6 was typical of this success (Fig. 6). Results from this spread list showed that F5 wobbled a bit more over the left side of the graph but reached accuracy with fewer ecoregions

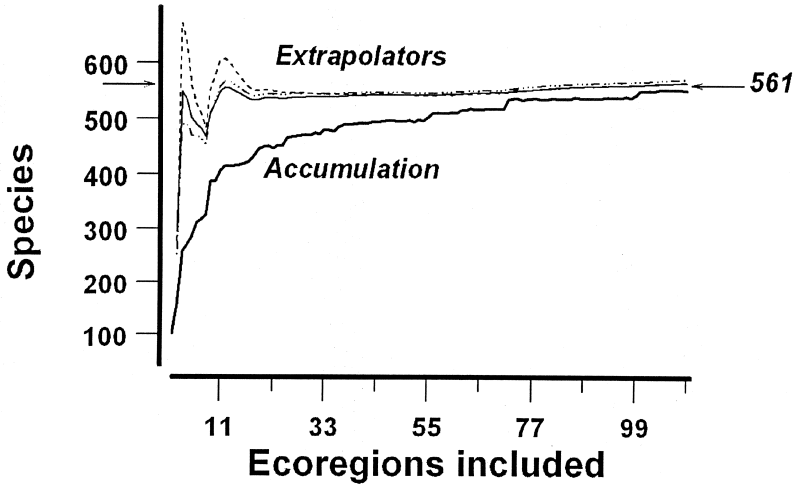

Figure 6. Extrapolation estimates from ecoregion spread list W6. The upwardly curving line exhibits the actual accumulation of species as ecoregions are added in the fixed order of W6. The other three lines are sequential estimates of total diversity generated by F3, F5, and Michaelis-Menten (MM). Each estimate is plotted over the number of ecoregions on which it is based. The estimators yielded similar results, but their trace can be most easily distinguished over two x *values:* x = 5 *(F5 = 559; MM = 501; F3 = 470) and* x = 12 *(F5 = 607; MM = 555; F3 = 566).*

Table 2. Extrapolation estimates of species diversity with 11 or 22 ecoregions chosen in fixed, hyperdispersed order ("spread lists") from the full set of 110.*

	Includes first 11 ecoregions				*Includes first 22 ecoregions*			
	F6	*F5*	*F3*	*MM*	*F6*	*F5*	*F3*	*MM*
Minimum	400.0	408.1	407.5	419.7	444.3	459.8	463.7	472.4
Maximum	2523.7	805.4	650.4	638.7	1072.1	636.1	617.6	600.4
Mean	804.0	530.5	500.7	504.9	660.8	546.6	539.3	533.7
Median	602.6	511.8	490.1	498.0	617.5	549.8	544.3	537.1
SD	536.3	100.0	68.0	59.0	164.4	48.6	43.0	35.7

**We ran 10 replicates of each of the 21 spread lists and then took statistics on the pool of their estimates. A perfect estimate remained 561 species. F5, F3, and MM all performed well but exhibited a more negative bias than the estimates from random trials.*

and held that accuracy as ecoregions accumulated. With all 110 ecoregions in the sample, F5 = 566, F3 = 571, and MM = 565. F6, the least successful extrapolation estimator, does not appear in Figs. 5 or 6. It produced overestimates as high as 3112 species on the left side and ended with an estimate of 590 species with all 110 ecoregions accounted for.

Because each spread list generates a separate set of results, spread lists produce estimates of variability with no special extra work. Their estimates of variability were much less variable then the estimates derived from random lists (Table 2). This reduction in variability probably comes from the nonrandom rules used to generate spread lists. However, the mean and median estimates of F3 and MM from spread lists had a greater negative bias than those from random lists. And the mean and median estimates of F5 from spread lists had a negative bias, whereas those from random lists appeared to show either no bias or a positive bias. In other words, spread lists did not generate estimates as close to 561 as random lists did. This negative bias did not characterize estimates from F6, but F6 had a much greater variability than the others and its estimates were not good enough to merit further mention.

Results from Combining Methods

To obtain the results reported above, we analyzed each of the 21 spread lists separately but in strict order. Species from the second region of each list were added to those of the first, then those of the third, and so forth. This inevitably led to considerable irregularity in the accumulation curves. They are much less smooth than those generated by the random-order method. (Fig. 6 shows an example.) Because of the strict order, our replications could not smooth the accumulation curves much, and we reported results from only 10 runs. We did wonder, however, whether we could improve the estimates from spread lists and further reduce their variability by combining the realism of the spread tactic with the smoothing power of random sampling.

Because random accumulation yielded considerable accuracy with only 10% of the ecoregions, we smoothed using the first 10% of the ecoregions in each spread list. We extracted the species occurrence records of the first 11 ecoregions in each of the 21 spread lists. Then we subjected each of these truncated lists to random analysis. This is quite a natural way to proceed. One may imagine that the explorer, having completed the first 11 ecoregion surveys, analyzes the results with the random tactic.

The first 11 ecoregions of a spread list can be listed in 11! orders. Each random run selects one of these orders. We executed 50 runs per list, which greatly smoothed the 21 raw accumulation curves (Fig. 7 shows two examples). The smoothing reduced the variance of all three estimators (Table 3). (The significance of this reduction is not at issue; we know a priori why it has to happen.) Although the reduction is small, it comes at no extra analysis or surveying cost.

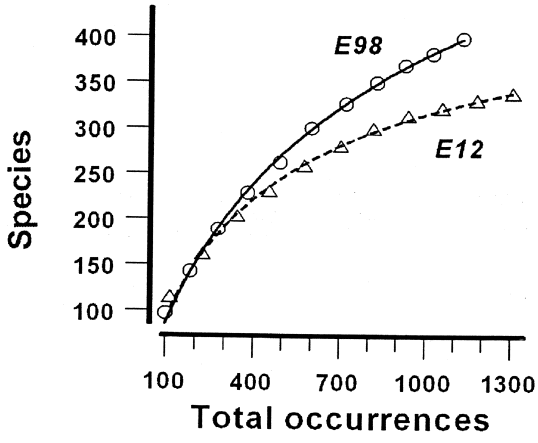

Figure 7. Estimating diversity from the first 11 ecoregions of two spread lists, E12 and E98. One occurrence is a single species in a single ecoregion. We ran each list 50 times, shuffling its ecoregion order to smooth the accumulation curves. Then we fit F5 to obtain their asymptotes. E12 yielded a low estimate, 549 species, and E98 yielded an unusually high estimate of 714. Symbols show the average total number of occurrences at each step. Lines show the F5 regressions.

Table 3. Extrapolation estimates of species diversity with ecoregions limited to the first 11 of each of the 21 "spread-lists."*

	Species in sample	*F5*	*F3*	*MM*
Minimum	326.3	448.7	431.5	436.2
Maximum	425.8	727.3	625.1	613.2
Median	392.3	574.2	540.5	534.6
Mean	386.9	582.3	537.0	533.0
SD	26.4	68.9	50.8	45.8
95% confidence interval		31.4	23.1	20.8

**We ran 50 replicates of each of the 21 spread lists and then calculated statistics on the pool of their 21 average estimates. In each of the 1050 analysis runs, ecoregion order was shuffled randomly. A perfect estimate remained 561 species. The median estimates of F5, F3, and MM were all quite close to 561, despite the limited sample. F5 tended to overestimate diversity. MM and F3 tended to underestimate it.*

More important than the smoothing, MM, F3, and F5 all now produced reasonably accurate average estimates of total diversity (including the species in the unsampled 99 ecoregions) when the random tactic operated on the occurrence records of the first 10% of the ecoregions in each spread list. F3 and MM tended to underestimate diversity a bit and F5 to overestimate it (Table 3).

Discussion

Can estimators of species diversity that compensate for small sample sizes also compensate for incomplete exploration of habitat types? We put this question to six estimators of species diversity. In particular, we tested whether they could produce worthwhile estimates of total butterfly species diversity in a continental-scale region (Canada and the United States).

Those of the six estimators that operate by sampling species abundance distributions (Burnham and Overton's jackknife and Lee and Chao's ICE) failed our tests. The other four operate by extrapolating the asymptote of a sequence of diversity accumulations. Three of these—MM (based on a simple Michaelis-Menten formula) and F3 and F5 (based on a new family of asymptotic formulas) succeeded. The remaining extrapolation formula, F6, based on the very same family, failed.

Provided with only 10% (i.e., 11) of the species lists from ecoregions, the three successful extrapolation estimators—MM, F3, and F5—did a remarkable job. Using a combination of tactics that mimics how a real investigator might work, they produced median estimates within 4.7%, 3.7%, and 2.4%, respectively, of the actual value (561 species). The best of these, F5, returned an estimate between 551 and 614 species in 95% of cases. For F3 the confidence interval was 514–560, and for MM it was 512–554.

The tactics of sequential exploration proved important. The investigator who begins at one place and gradually explores from there is likely to fail, even using the best extrapolation estimator. Exploration sites may be added gradually, but their locations must not be clustered. They should at least be chosen randomly within the total area under investigation. Better even than random siting was the spread tactic, in which locations were added so that exploration sites were always spread rather uniformly through the total area.

The residual variation in estimated diversity arose from the variation in total species accumulated by the first r ecoregions in a spread list. Total species accumulated (S_{obs}) correlated positively with diversity estimates. (Regression results: F3 prediction for the first 11 ecoregions of the 21 spread lists regressed on S_{obs}, $p = 10^{-5}$, $R^2 = 0.65$; F5 prediction on $S_{obs,}$ $p < 10^{-3}$, $R^2 = 0.46$; MM prediction on $S_{obs,}$ $p = 10^{-5}$, $R^2 = 0.64$.) We found no way to eliminate that correlation, not even by increasing the number of ecoregions included. Nevertheless, the extrapolation estimators we studied gave a good empirical picture of the value of true S.

The primary difference between the three successful estimators and the fourth extrapolation estimator (F6) is that F6 has two parameters of curvature (a and q) instead of only one. Despite that extra parameter of curvature, F6 failed to extrapolate to the true diversity, whereas the other three extrapolators—each with only one parameter of curvature—succeeded. We do not understand why F6 did not do as well as the other three.

We also cannot explain why MM did succeed. The Michaelis-Menten formula (on which MM is based) has a built-in error. It does not go through the point (1,1), a point that characterizes every diversity sample with a single individual. Perhaps this error should have been a heavy burden for MM, but it was not. Is MM's success limited to large ecoregions containing many species? Will MM's intrinsic error be more consequential when estimates are being made at smaller scales with relatively small amounts of information?

We believe that we do understand the failures of the jackknife and ICE estimators. These two were designed to overcome sample-size inadequacies and to reveal how many species are present in habitats actually sampled. Our results do not challenge their success at doing this job. But neither one was designed for extrapolation. They operate only on the results obtained from a particular subset of the total data set. They pay no attention to the pathway along which that subset was accumulated. But that pathway is what must contain the information needed to reveal the asymptote of diversity. And that pathway is the very pathway that the extrapolation formulas are meant to fit.

F3 and F5 are hybrids. They combine deductive reasoning with empiricism. A priori reasoning tells us that a

successful formula must begin at the point (1,1) and rise monotonically toward an asymptote. Data tell us that the second derivative of any successful formula must be negative throughout. (Some formulas conforming to Eq. 2 produce curve segments with positive second derivatives; we did not include them in this report.) And, of course, the data are the ultimate judges of F3 and F5. Except for its failure to traverse the point (1,1), MM also combines these same deductive and empirical components.

In arriving at F3 and F5, we included no assumptions except to begin at the point (1,1) and rise monotonically toward an asymptote. In particular, neither formula assumes a specific distribution of abundances (such as log series or lognormal), which we believe is a strength. Just as many processes produce linear relationships (thus fitting the general skeletal equation, $y = mx + b$), so F3 and F5 will fit many curves that rise to an asymptote from the point (1,1) (as will F6, although it served poorly for diversity extrapolation).

No formula passes muster merely by making good theoretical sense. The data must also support it. The central role of the data in arriving at our conclusions is incontrovertible. Having no complete deductive scheme to predict our extrapolation formulas, we must admit that their excellent fit to the butterflies of North America might turn out to be unusual. More real data from other taxa and other provinces will prove essential to acceptance of the methods. The present results, however, seem promising.

The butterfly data come from the large scale of whole ecoregions. Moreover, they are quite unusually complete. Does successful extrapolation depend on large scale and fairly complete knowledge of some of its components? We have no answer to the question of scale, but complete knowledge of components may not be essential. Bias-reducing estimators such as ICE and the jackknife efficaciously tell us how many species reside in a given patch of space-time. A hyperdispersed series of such patches could be operated on by such estimators to produce good estimates of diversity within each patch. If we could extract from such data an estimate of species overlap from ecoregion to ecoregion, then, operating on a growing set of patches, the present extrapolation formulas could produce an estimate of the asymptote. The extrapolation ought to be valid for the whole area within which the samples have been selected.

We used only geographical location to organize the (hyperdispersed) spread lists. Would it have been better to rely on the biologically relevant properties of a place, for example, its temperature and rainfall? We doubt it. One can usually get such data for large scales everywhere, but not for microscales. If we had used a biologically meaningful set of variables to sort ecoregions, we might have raised an impediment to use of the technique at microscales. Thus, we are particularly encouraged that the simple tactic of spreading the sampling locations worked well.

The practical application of the method requires an answer to at least one other question. How many subsamples are enough? We found that a 10% sample was enough, but that answer may not always be correct. In practice, moreover, it may be difficult to tell when 10% is reached (especially at smaller scales). We are actively working on developing an internal statistic to answer this question. A promising candidate involves the ratio of accumulated species to estimated species.

A successful extrapolation estimator will have many uses. It will greatly reduce the time and resources required to obtain reliable diversity totals (Heywood 1995). It will allow diversity to be estimated at fine temporal scales so that conservation biologists can better track its dynamics (Devries et al. 1999). It should manifestly improve the ability of paleobiologists to overcome the incompleteness of the fossil record and so better allow them to discern any patterns of change in diversity in the history of life (Lee 1997). Finally, it will at last enable ecologists to estimate worldwide species diversity (Gaston & Hudson 1994; Rosenzweig 1995).

Acknowledgments

A portion of this paper is derived from the University of Arizona senior honors thesis of J. G. Cox. E. Siemann furnished the data for Fig. 1. We thank H. Possingham, G. Meffe and E. Main for suggestions that improved the manuscript.

Literature Cited

Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology **60:**927–936.

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, editors. Forest biodiversity research, monitoring and modeling: conceptual background and old world case studies. Volume **20**. United Nations Environmental, Scientific, and Cultural Organization, Paris, and Parthenon Publication, New York.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London Series B **345:**101–118.

Devries, P. J., T. R. Walla, and H. F. Greeney. 1999. Species diversity in spatial and temporal dimensions of fruit-feeding butterflies from two Ecuadorian rainforests. Biological Journal of the Linnean Society **68:**333–353.

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.

Gaston, K. J., and E. Hudson. 1994. Regional patterns of diversity and estimates of global insect species richness. Biodiversity and Conservation **3:**493–500.

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

Heywood, V. H. 1995. Global biodiversity assessment. United Nations Environment Programme, Cambridge, United Kingdom.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology **54:**427–431.

Holdridge, L. R., W. C. Grenke, W. H. Hatheway, T. Liang, and J. A. Tosi. 1971. Forest environments in tropical life zones. Pergamon Press, Oxford, United Kingdom.

Lee, M. S. Y. 1997. Documenting present and past biodiversity: conservation biology meets palaeontology. Trends in Ecology & Evolution **12:**132–133.

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

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

Poulin, R. 1998. Comparison of three estimators of species richness in parasite component communities. Journal of Parasitology **84:** 485–490.

Ricketts, T. H., E. Dinerstein, D. M. Olson, C. Loucks, W. Eichbaum, K. Kavanagh, P. Hedao, P. Hurley, K. M. Carney, R. Abel, and S. Walters. 1999. Terrestrial ecoregions of North America: a conservation assessment. Island Press, Washington, D.C.

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

Siemann, E., D. Tilman, and J. Haarstad. 1996. Insect species diversity, abundance and body size relationships. Nature **380:**704–706.

Smith, B., and J. B. Wilson. 1996. A consumer's guide to evenness indices. Oikos **76:**70–82.

Soberón, J., and J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology **7:**480–488.

Turner, W., W. A. Leitner, and M. L. Rosenzweig. 2000. Ws2m.exe. University of Arizona, Tucson. Available at http://eebweb.arizona.edu/diversity (accessed 31 March 2003).