On Species-Area Relations

Fangliang He; Pierre Legendre

American Naturalist, Volume 148, Issue 4 (Oct., 1996), 719-737.

ON SPECIES-AREA RELATIONS

Species-area relationships have been of interest in ecology for a long time (de Candolle 1855; Jaccard 1902, 1908). They have been called "one of community ecology's few genuine laws" (Schoener 1976, p. 629). Plotting number of species (S) against sampling area (A), for a series of samples of increasing sizes, yields a monotonically increasing curve whose slope is steep at first but gradually becomes nearly flat. The shape of such curves has been used to help determine the area required to obtain an adequate sampling of the species in a particular community—the "minimal area" concept (Goodall 1952; Hopkins 1957; Cain and Castro 1959; Barkman 1989), to characterize community structure (Fisher et al. 1943; Goodall 1952; Preston 1962; May 1975), to estimate species richness (Evans et al. 1955; Kilburn 1966; Hubbell and Foster 1983; Palmer 1990; Baltanás 1992; Grassle and Maciolek 1992), to measure the effect of disturbance on communities (Lawrey 1991), and to define the appropriate size of reserves and natural areas in conservation biology (MacArthur and Wilson 1967; Soulé et al. 1979; Williamson 1981).

A large number of natural communities have been investigated, especially by plant ecologists. They found that different communities may exhibit different types of species-area relations. Among them, three expressions are most widely used: exponential curve (Gleason 1922, 1925),

$$S = Z \ln(A) + C; \tag{1}$$

power curve (Arrhenius 1921, 1923a, 1923b),

$$S = C A^Z; \tag{2}$$

and logistic curve (Archibald 1949),

$$S = \frac{B}{C + A^{-Z}}, \tag{3}$$

where S is the number of species, A is the area, and B, C, and Z are constants by conventional notation.

These models were first proposed as the result of empirical observations, and they have competed with one another in ecology. There has been wide disagreement among researchers on three topics: what is the best model, how to explain the best model in ecological or biological terms, and how to estimate and interpret the parameters of species-area models.

In the present study, our main purposes are to attempt to unify the above three species-area curves using a general model, to relate different curves to different

sampling situations, and to argue that the logistic curve is an appropriate model to estimate asymptotic species richness in a community or a region.

The three species-area models described above are special cases of a general equation. We can safely assume that the derivative of the number of species (S) with respect to sampling area (A)—that is, the rate at which new species are found when the area increases dA—is the ratio of some function of the number of species to the sampled area:

$$\frac{dS}{dA} = \frac{f(S)}{A} . \tag{4}$$

We assume that $f(S)$ has three continuous derivatives on $S \in (0, \infty)$. Although we do not know at this point what $f(S)$ looks like, we can expand it into a Taylor series about $S = 0$, that is,

$$f(S) = f(0) + f'(0)S + \frac{1}{2}f''(0)S^2 + o(S^3) , \tag{5}$$

where $o(S^3)$ is the remainder of the expansion. Replacing $f(0)$, $f'(0)$, and $f''(0)/2$ by constants α, β, and γ, we obtain

$$f(S) = \alpha + \beta S + \gamma S^2 + o(S^3). \tag{6}$$

The exponential, power, and logistic species-area models can be derived from equation (6), truncating the terms of order higher than the first, second, and third, respectively.

Exponential Model

Retaining only the first term of equation (6), equation (4) becomes

$$\frac{dS}{dA} = \frac{\alpha}{A} . \tag{7}$$

Rearranging equation (7), the integral expression is as follows:

$$\int dS = \int \alpha \, \frac{dA}{A} . \tag{8}$$

The integral of equation (8) is

$$S = \alpha \ln A + c , \tag{9}$$

where c is an integral constant. We easily obtain the exponential model (1) by making $Z = \alpha$ and $C = c$.

Power Model

Retaining the first two terms of equation (6), equation (4) becomes

$$\frac{dS}{dA} = \frac{\alpha + \beta S}{A} . \tag{10}$$

Rearranging equation (10), we obtain the following integral expression:

$$\int \frac{dS}{\alpha + \beta\,S} = \int \frac{dA}{A}\,. \tag{11}$$

After integrating and a few steps of rearrangement, we obtain

$$\frac{1}{\beta} \ln(\alpha + \beta\,S) = \ln A + c\,, \tag{12}$$

where c is a constant. Rearranging the above equation produces

$$\alpha + \beta\,S = e^{\beta c}\,A^{\beta}\,. \tag{13}$$

Imposing the constraint that the number of species $S = 0$ if the sampling area $A = 0$, which is another fair assumption, leads to $\alpha = 0$ in equation (13). The consequence is that in equation (10), we only need to include the second term of equation (6). Thus, equation (13) reduces to

$$S = \frac{e^{\beta c}}{\beta}\,A^{\beta}\,. \tag{14}$$

Taking $Z = \beta$ and $C = e^{\beta c}/\beta$, we obtain the power model (2).

Logistic Model

   This time, we use the first three terms of equation (6). After derivations similar to those above, we can obtain the logistic model of equation (3). Here $\alpha = 0$ is obtained by assuming that $S = 0$ if $A = 0$, under the condition $\gamma \neq 0$. The derivative is the following:

$$\frac{dS}{dA} = \frac{\beta\,S + \gamma S^{2}}{A}\,. \tag{15}$$

By integrating equation (15) and rearranging the parameters, we obtain the logistic model of equation (3), with $B = \beta e^{\beta c}$, $C = -\gamma e^{\beta c}$, and $Z = \beta$. Because the number of species available for colonization in any biogeographical region or geographical unit of the planet is limited, the species-area curve has to be asymptotic to some upper "number of species" bound. So, when the logistic function is fitted to any real species-area data set, γ will automatically take up a negative value, with the consequence that C in equation (3) is positive; the ratio $B/C = \beta/(-\gamma)$ gives the maximum number of species (asymptote of the curve) in the community.

   In summary, equation (4) dynamically describes the relationship between the number of species in a community and the area they occupy. The exponential, power, and logistic models are obtained by using the first one, the first two, or the first three terms of equation (6), respectively. The merit of this model is that it unifies the species-area models and makes them comparable under a single mathematical framework. The links among these models are obvious from equations (7), (10), and (15). Equation (15) embodies the effect of equation (10), while equations (10) and (15) possess the effect of equation (7) under the condition
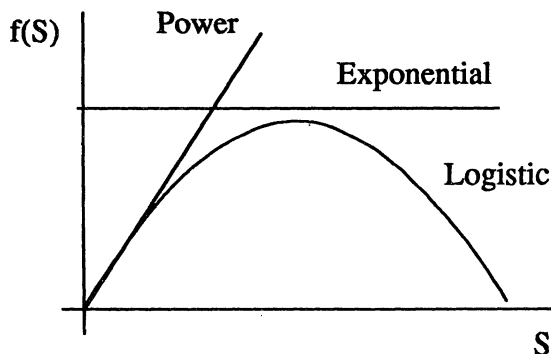
Fig. 1.—Illustration of function $f(S)$ for the exponential, power, and logistic species-area models.

that $S = 0$ if $A = 0$. The three models of $f(S)$ as a function of S are shown in figure 1. When S is small (approaching zero), equations (10) and (15) converge.

In equation (4), $f(S)$ is a function of the number of species, describing how this number responds to the increasing size of the sampling area. It is very much like the response of population growth to available resources in the environment, in Malthusian and logistic models (the former for unlimited, the latter for limited resources). Here the area is an environmental resource available for occupation by the species present in the zone. In a very small area, a small rise in sampling area approximately leads to a linear increase in the number of species; increasing the sampling area further, the number of species rises in an accelerated way following the power relation; with an even larger increase in area, the increase in number of species slows down and finally levels off.

In a given time, the number of species in a community can be considered constant. Based on the properties of equation (4) and on our understanding of $f(S)$, if that community is only sampled across a small area, then the species-area relation should follow the exponential model; for a sampling area of intermediate size, the relation should follow the power model; and for a large sampling area, the logistic model becomes suitable. The "small," "intermediate," and "large" sampling areas are relative to the size of the community under study. These species-area relations are more clearly illustrated in figure 2.

This line of reasoning is supported by the findings of several authors, who note that species-area relations often possess an inflection point (in S–ln A space), corresponding to the point where a large proportion of the community area has been sampled; as a consequence, these species-area curves are sigmoid in shape (Archibald 1949; Vestal 1949; Niering 1963; Whitehead and Jones 1969; Abbott 1973; Lassen 1975; He et al. 1996). Williams (1943) found that samples from small areas had species-area relations in the form of the exponential model, while the power model was more appropriate for intermediate sampling areas. Kilburn (1963) also found that the power model was best fitted to samples from a relatively
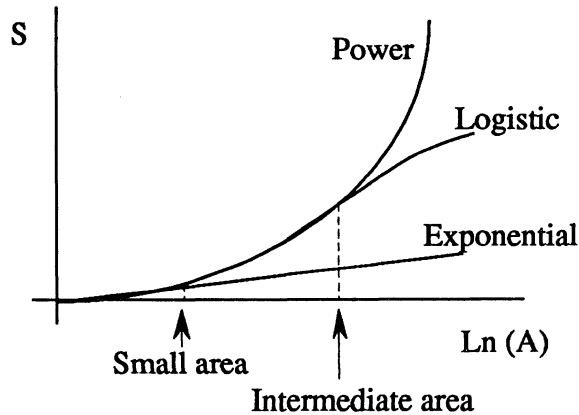
FIG. 2.—Illustration of the three species-area models, as functions of ln(A) in abscissa. Since the community species-area curve is logistic, the picture shows that the exponential model is only valid for species-area data from a small sampling area relative to the size of the community. The power model can be used either for small or intermediate sampling areas; but if the community area is well sampled, the logistic model should be used.

small sampling area, and he suggests that the logistic model is best for data from larger sampling areas.

STUDY CASES

Statistical Criteria to Select the Best Model

Connor and McCoy (1979) suggest that the best-fit model should be selected empirically, using statistical criteria. They discriminated among competing species-area models by checking for no lack of fit in the regression residuals or by the size of the correlation coefficient, and by retaining all models that fit the data well. Though Sugihara (1981) suggests that the best model should be judged by assessment of the lack of fit, not by absence of lack of fit, Connor et al. (1983) argue that comparing regression or residual sums of squares would be an appropriate means to discriminate among models. In practice, strict assessment of lack of fit is virtually impossible because multiple observations (i.e., replicate measurements) of species richness within a fixed sampling quadrat are hardly ever available (for the method to assess lack of fit of a model, see Jobson 1991, pp. 135–137). In any case, though Connor and McCoy's (1979) criteria to select a best model are useful and informative, if there are different numbers of parameters between models, a lower residual sums of squares or a higher correlation coefficient does not guarantee the best model, because the extra parameters alone may contribute to these statistics. A formal test of significance must be carried out.

In this note, we will compute all three models and select the best-fitting one, using as best-fit criterion the probability associated with the usual F ratio (regres-

sion mean square over the error mean square) of the ANOVA table of the regression model. The best model is the one with the lowest P value (probability under the null hypothesis). We also computed the adjusted coefficient of determination (Ra^2), as used by Boecklen and Gotelli (1984) and Loehle (1990) for species-area model fitting. This coefficient is calculated as

$$Ra^2 = 1 - \frac{(n-1)\,RSS}{(n-k)\,TSS}, \tag{16}$$

where RSS is the residual sum of squares, TSS is the total sum of squares, n is the number of samples, and k is the number of parameters in a model. This coefficient is more suitable than the usual coefficient of determination R^2 in that it takes into account the respective numbers of degrees of freedom of the numerator and denominator. The purpose of this correction is to produce a statistic suitable for comparing regression equations fitted to different data sets, with different numbers of samples and numbers of independent variables.

All the calculations were made using the nonlinear regression program NLIN of SAS version 6.08 (SAS 1989); the iterative method was Marquardt. Probabilities were calculated using a Fortran subprogram for computing the probability of an F ratio, adapted from Dorrer (1976).

A Tropical Rain Forest Data Set (Plant Community)

A data set describing the tree community structure of a lowland tropical rain forest, in the Pasoh Forest Reserve of Malaysia (called the Pasoh forest hereafter; see Kochummen et al. 1991), was used to fit equations (1), (2), and (3), using different sampling areas (A). The entire study area is 50 ha (500 × 1,000 m). The survey consisted of enumerating all free-standing trees and shrubs at least 1 cm in diameter at breast height (dbh), positioning each one by geographical coordinates on a reference map, and identifying it to species. The diversity of the plot was quite high: there were 334,077 trees, belonging to 825 species. To prevent overlap in sample areas, which would make the samples interdependent, we proceeded as follows. Using a computerized database of the Pasoh forest, the study area was divided into two equal parts, and one of them was selected at random as our largest-size sample; the remaining part was then divided again into two equal parts, and one of them was randomly selected; the procedure was repeated 17 times. At the last division (into quadrats of 3.81 m²), only one quadrat was selected. The objective was to retain only one quadrat of each size for model fitting. The raw data and the codes of the sampled quadrats are shown in table 1.

First, the exponential and power models were fitted to the data from the smallest sampling areas only: Ib, II, III, and IV (fig. 3a). Because the species-area relation for the small samples did not display an inflection point, the logistic model could not be fitted. The observed species-area curve and the fitted logistic curve for all sampling areas (3.81–250,000 m²) are also shown in all panels of figure 3 for reference. Although both the exponential and power models satisfactorily describe the species-area data for small samples, the exponential model gives better-fitting results (P value and Ra^2; table 2). The fitted equations are

TABLE 1

RAW DATA AND PREDICTED NUMBERS OF SPECIES FOR THE PASOH FOREST OF MALAYSIA, USING THE
EXPONENTIAL, POWER, AND LOGISTIC MODELS

CODE	AREA (m²)	OBSERVED NUMBER OF SPECIES	PREDICTIONS USING THE FOLLOWING MODELS		
			Exponential	Power	Logistic
Ia	3.81	3	———	57.7	9.9
Ib	3.81	3	———	57.7	9.9
II	7.63	3	———	68.5	15.0
III	15.26	12	———	81.2	22.7
IV	30.52	13	28.9	96.3	34.1
V	61.04	31	83.1	114.2	51.0
VI	122.07	70	137.2	135.4	75.3
VII	244.14	112	191.3	160.6	109.6
VIII	488.28	134	245.5	190.5	156.3
IX	976.56	236	299.6	225.9	217.0
X	1,953.12	308	353.8	267.9	291.1
XI	3,906.25	400	407.9	317.7	375.0
XII	7,812.5	471	462.0	376.8	462.3
XIII	15,625	541	516.2	446.8	545.7
XIV	31,250	593	570.3	529.9	618.9
XV	62,500	631	624.5	628.4	678.6
XVI	125,000	752	678.6	745.3	724.3
XVII	250,000	783	732.7	883.8	757.9

NOTE.—Sampling size doubles from one quadrat to the next but is subjected to rounding error. The sum of these areas is the whole size of the Pasoh plot (500,000 m²). Code Ia was excluded from model fitting in order to keep only one quadrat of each size in the calculations. A dash represents negative numbers of species predicted by the exponential model.

exponential model, $S = 5.623 \ln(A) - 5.623$; power model, $S = 1.313 A^{0.694}$. Using the fitted exponential model equation, the estimated number of species for the 3.81-m² sampling area is 1.9. Using the power model equation, it is 3.3; by comparison, the observed value for that sampling area is 3.0. For small sampling areas, the exponential model systematically produces large deviations from the observed data. For very small sampling areas, it predicts negative values that are meaningless.

Nine sampling quadrats from Ib to IX were then analyzed together (table 1). The fitted curves for the exponential and power models are presented in figure 3b for that range of samples; the power model emerges as a better-fitting description of the observations than the exponential (table 2). The fitted equations are exponential model, $S = 46.528 \ln(A) - 115.217$; power model, $S = 4.011 A^{0.578}$. Finally, the same evaluation of the three models was carried out with all samples from Ib to XVII (table 1). Although the iterative estimation process converges for all three models, the logistic model provides the best-fitting results (table 2). The fitted equations are exponential model, $S = 78.108 \ln(A) - 238.103$; power model, $S = 41.546 A^{0.246}$; and logistic model, $S = 4.438/(0.00534 + A^{-0.609})$.

Lake Islands Data Set (Bird Community)

On islands of the Pymatuning Lake, at the Pennsylvania-Ohio border, Coleman et al. (1982) collected a complete breeding bird data set in 1978 and 1979; hereaf-
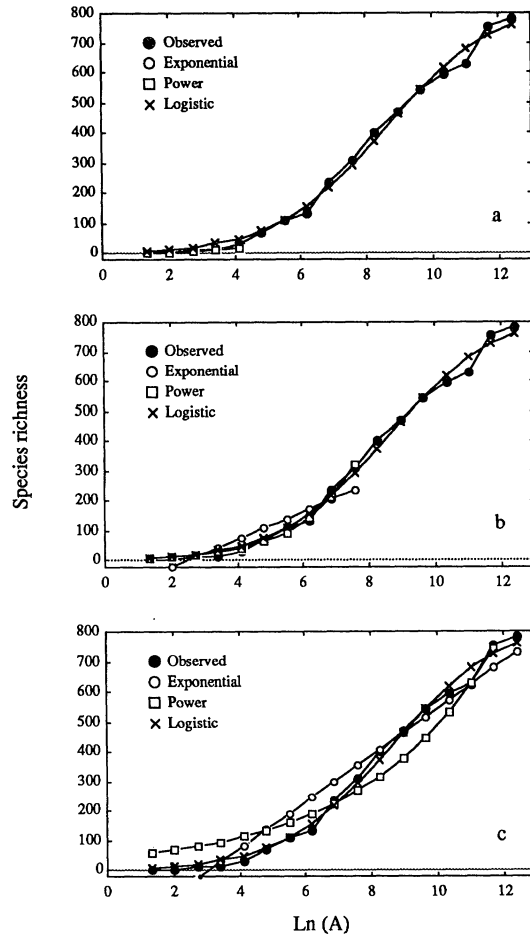
FIG. 3.—Species-area curves for the Pasoh forest data, Malaysia. The observed data and the fitted logistic curve for the whole sampling area are presented in all three panels for reference. *a,* Small sampling areas, 3.81–30.52 m^2: the species-area data are best fitted by the exponential model (the fitted curve is hidden by the power curve). *b,* Small to intermediate-size sampling areas, 3.81–976.56 m^2: the power model is better than the exponential. *c,* Small to large sampling areas, 3.81–250,000 m^2: the species-area data are best fitted by the logistic model.

ter, the two data sets are called Cole78 and Cole79. These authors found that the species-area relation across the islands was best explained by the random placement model, which is derived from the hypothesis that all individuals in an area are located at random. This model was superior to the exponential and power models for this particular data set (Coleman et al. 1982). We reanalyzed this data set using the exponential, power, and logistic models and compared the results using the "root mean-square deviation" (Δ), the same criterion used by Coleman et al. (1982); this statistic is the square root of the RSS divided by the number of

TABLE 2

ADJUSTMENT OF THREE SPECIES-AREA CURVES TO THE PASOH FOREST DATA OF MALAYSIA

Sampling Area*	Exponential	Power	Logistic
Small:			
RSS	14.72	17.80	———
F ratio	21.49	17.59	———
df	2,2	2,2	———
Probability	.0455†	.0538	———
R_a^2	.933	.919	———
Small to intermediate:			
RSS	16,421.05	1,572.71	———
F ratio	41.61	472.25	———
df	2,8	2,8	———
Probability	5.914×10^{-5}	4.976×10^{-9}†	———
R_a^2	.901	.991	———
Small to large:			
RSS	68,399.37	76,163.18	7,375.80
F ratio	298.44	267.26	1,760.68
df	2,15	2,15	3,14
Probability	8.330×10^{-13}	1.866×10^{-12}	$<10^{-30}$†
R_a^2	.974	.971	.997

NOTE.—A dash represents negative numbers of species predicted by the exponential model.

* Small sampling areas correspond to codes Ib–IV, 3.81–30.52 m² (see table 1); small to intermediate sampling areas, codes Ib–IX, 3.81–976.56 m²; and small to large sampling areas, codes Ib–XVII, 3.81–250,000 m².

† The best-fitting model (lowest probability of the data under the null hypothesis).

samples. Coleman et al. (1982) evaluated the exponential and the power models using linear regression analysis; we used nonlinear regression analysis instead and obtained somewhat different results, showing that different regression methods can give rise to different estimates (table 3). Although the conclusion of Coleman et al. (1982) was not altered (i.e., the random placement is superior to the exponential and power models), all of the nonlinear evaluations had smaller root mean squares than their linear regression counterparts (table 3). On the other hand, we found that the logistic model is as good as (Cole78) or superior (Cole79) to the random placement model.

Having shown that the logistic curve is an appropriate model, we fitted the exponential, power, and logistic models to different island sizes: small islands only, small to intermediate islands, and small to large islands. We obtained similar results as in the tropical rain forest community of the previous section: the exponential model best fits small samples, the power model best fits small to intermediate samples, while the logistic model is the best for small to large samples (table 4).

DISCUSSION

Models and Hypotheses

The study of species-area relationships has long focused on the shape of the relation, its interpretation, and the mechanisms explaining the proposed form

THE AMERICAN NATURALIST

TABLE 3

COMPARISON OF THE EXPONENTIAL, POWER, LOGISTIC, AND RANDOM PLACEMENT MODELS FOR TWO
RESIDING BIRD SPECIES-AREA DATA SETS (COLE78 AND COLE79) FROM THE ISLANDS OF
PYMATUNING LAKE (COLEMAN ET AL. 1982)

PARAMETERS	COLE78	COLE79
	15 Samples (Islands)	27 Samples
Exponential model:		
C	6.506 (6.549)	8.333 (8.333)
Z	14.351 (14.453)	12.742 (12.743)
$\Delta_{(E)}$	2.38 (3.9)	3.43 (3.4)
Power model:		
C	6.285 (4.298)	6.476 (4.993)
Z	.425 (.600)	.428 (.543)
$\Delta_{(P)}$	3.05 (5.8)	3.24 (4.1)
Logistic model:		
B	5.537	6.776
C	.142	.184
Z	.916	.998
$\Delta_{(L)}$	1.67	2.02
Random placement $\Delta_{(R)}$	(1.7)	(2.2)

NOTE.—Terms with Δ the root mean square of deviation. Parameters are defined in equations (1), (2), and (3) and were evaluated by nonlinear regression analysis. To remain comparable with Coleman et al. (1982), calculations were done using base-10 logarithms wherever applicable. The values in parentheses have been estimated by Coleman et al. (1982) using linear regression analysis.

(Connor and McCoy 1979; McGuinness 1984). There have been in general three approaches to evaluate species-area curves. The first approach proposes and evaluates species-area relations by empirically fitting existing equations to field observations (Arrhenius 1921; Gleason 1922; Archibald 1949; Hopkins 1955; Kilburn 1963, 1966). The second approach attempts mathematically to derive species-area curves on the basis of observed species-abundance relations and the assumption of a linear relationship between number of individuals and area (Preston 1960, 1962; Williams 1964; May 1975; Engen 1977; Wright 1988). The third approach justifies the use of a given species-area curve from some ecological (or statistical) mechanisms and hypotheses (Arrhenius 1921; Preston 1962; MacArthur and Wilson 1963, 1967; Williams 1964; Coleman 1981). The last approach is the most interesting in ecology because it invokes mechanistic explanations. Unfortunately, these explanations may not generate unique predictions; for example, both the equilibrium hypothesis and the habitat diversity hypothesis can explain the power model (Williams 1943; MacArthur and Wilson 1963). Connor and McCoy (1979) suggest that the contribution of a specific ecological mechanism to an observed species-area relation should only be assessed by experimentation. This is what Simberloff and Wilson (1969) did for the recolonization of mangrove islands that had been experimentally defaunated. In our view, however, even with field experiments, the problem cannot fully be resolved, because although an experiment can support a given species-area relation, it cannot falsify alternatives.

TABLE 4

ADJUSTMENT OF THREE SPECIES-AREA MODELS TO THE TWO SPECIES-AREA DATA SETS (COLE 78 AND COLE79) FROM THE ISLANDS OF PYMATUNING LAKE (COLEMAN ET AL. 1982)

	COLE78			COLE79		
SAMPLE SIZE	Exponential	Power	Logistic	Exponential	Power	Logistic
Small samples:*						
RSS	12.33	13.71	—	36.71	36.91	—
F ratio	45.97	40.92	—	57.45	57.07	—
df	2,8	2,8	—	2,18	2,18	—
Probability	4.11×10^{-5}†	6.28×10^{-5}	—	1.53×10^{-8}†	1.62×10^{-8}	—
R_a^2	.910	.900		.857	.856	
Small to intermediate samples:‡						
RSS	71.83	42.10	41.32	297.00	65.46	59.60
F ratio	80.18	140.70	86.94	58.66	306.84	215.63
df	2,11	2,11	3,10	2,23	2,23	3,22
Probability	2.76×10^{-7}	1.46×10^{-8}†	1.83×10^{-7}	9.29×10^{-10}	$<10^{-30}$†	5.55×10^{-16}
R_a^2	.930	.959	.956	.829	.962	.964
Small to large samples:§						
RSS	85.02	139.68	41.66	317.14	283.30	109.75
F ratio	227.37	135.86	289.68	133.84	151.33	262.65
df	2,13	2,13	3,12	2,25	2,25	3,24
Probability	7.68×10^{-11}	1.94×10^{-9}	1.86×10^{-11}§	4.41×10^{-14}	1.08×10^{-9}	$<10^{-30}$§
R_a^2	.970	.951	.984	.911	.921	.968

NOTE.—A dash represents negative numbers of species predicted by the exponential model.
* For 1978, size was 0.27–1.44 ha; for 1979, 0.09–0.94 ha.
† The best-fitting model (lowest probability of the data under the null hypothesis).
‡ For 1978, size was 0.27–9.40 ha; for 1979, 0.09–9.40 ha.
§ For 1978, size was 0.27–69.44 ha; for 1979, 0.09–69.44 ha.

Williams (1943, 1964) proposes habitat diversity as a hypothesis explaining the correlation between species and area: with the increase of sampling area (A), from a local community to regional flora and to global biosphere, a greater diversity of environments is included, and so the number of species increases. In the real world, of course, the environment is never homogeneous, be it within or among communities. Even in the relatively homogeneous lowland tropical rain forest of Malaysia, heterogeneity is still detectable (He et al. 1994). In the general species-area model (eq. [4]), the derivative dS/dA describes how many new species dS are recruited when the area increases dA. With the increase of sampling area, heterogeneity inevitably increases, and with it the number of species. This process is not only valid for invertebrate species inhabiting a single plant population (Gunnill 1982) or a tropical rain forest (He et al. 1994), but also for islands (Abbott 1974; Werff 1983) and continental regions (Williams 1943, 1964). We have shown, however, that the change of the rate of dS/dA is not constant, and so a species-area relation takes different forms depending on the spatial scale of sampling. At a very local scale, the addition of new species is relatively slow within an area, probably because of a strong similarity (autocorrelation) of environmental conditions and species composition among neighboring sites. In these conditions, species-area relations are well described by the exponential model. Increasing the spatial scale, substantial heterogeneity is added as soon as we go beyond the range of action of the spatial autocorrelation of environmental conditions (Dutilleul and Legendre 1993). The number of new species increases more rapidly; consequently, the species-area relation is best described by the power model. Naturally, heterogeneity is not unlimited in a geographically defined community, ecosystem, or region (delimited by physical barriers or conceptually); neither is the number of species. If the sampling area keeps increasing, recruitment of new habitats is bound to slow down and finally cease; at the same time, the rate at which new species are found decreases until a different type of community is encountered; this process is best described by the logistic model.

Although the general model (eq. [4]) can reasonably be explained by the habitat diversity hypothesis, we do not limit the meaning of function $f(S)$ to mechanistic explanations. We believe this function is related to the spatial distribution of heterogeneity and species in a study area and to species-abundance relations, because the only term that can accommodate these factors in model (4) is $f(S)$. It is well known that different species-abundance relationships can give rise to different species-area curves (Williams 1943; Preston 1962; McGuinness 1984; Colwell and Coddington 1994); any spatial distribution of species that deviates from randomness would affect species-area curves according to the random placement hypothesis. It is not surprising that model (4) is also affected by these factors, although it is technically a challenge to incorporate them into $f(S)$. In a very general form, we can qualitatively express $f(S)$ as $f(S[\text{space, abundance}])$, considering spatial features and species abundances. Keeping this general form in mind, we can explain the effect of disturbances on species-area curves as a

TABLE 5

COMPARISON OF THE Z AND C VALUES OF THE POWER MODEL FOR DIFFERENT SAMPLING AREAS FROM
THE PASOH FOREST OF MALAYSIA (PASOH) AND THE 1978 AND 1979 DATA OF COLEMAN ET AL. (1982)

Sampling Area	Pasoh	Cole78	Cole79
Small:			
Z	.694	.573	.344
C	1.313	4.534	4.400
Small to intermediate:			
Z	.578	.653	.701
C	4.011	4.794	5.379
Small to large:			
Z	.246	.425	.429
C	41.546	6.285	6.476

NOTE.—Sampling areas are described in table 2 for the Pasoh Forest and in table 4 for the data of Coleman et al. (1982).

force modifying the spatial distribution of environmental heterogeneity, species distributions, and the abundance of species.

When the root mean-square deviations (Δ) for the logistic and random placement models are compared in table 3, it seems that the observed data can equally well be explained by the logistic rationale and the random placement hypothesis. These models gave similar results, although the logistic performs slightly better. The closeness between these two models may result from the bird species on the islands of the Pymatuning Lake being in fact randomly distributed. Were the species not randomly distributed, the performance of the two models would be much different. Once again, we turn to the Pasoh forest data to discriminate between these two models. The spatial distributions of all 825 species of the Pasoh forest have been thoroughly examined (He et al., in press). Excluding the species with an abundance of fewer than five individuals in the plot, 80.4% of the remaining 745 species presented aggregated distributions. Computing the root mean-square deviations (Δ) for the logistic and the random placement models for this data, we find substantial differences: $\Delta_{(L)} = 412.96$ for the logistic model, and $\Delta_{(R)} = 2,217.51$ for the random placement model. Therefore, we conclude that for a community in which species do not present random distribution (which is the picture most often encountered in nature), the logistic model is a better explanation than the random placement model. The two models converge only in the case of random spatial distributions.

Across a range of spatial scales, it is not only the species-area models that change but also the model parameters. For example, the Z value of the power model, which is possibly a parameter of interest in ecology, changes substantially for different sampling scales (table 5) for the Pasoh forest data and the two data sets of Coleman et al. (1982). At the whole-plot scale, the Z value for the Pasoh forest can be even lower than Preston's ideal value of 0.262; this is what was expected for "areas that are merely samples of larger areas" (Preston 1962, p. 191), such as island archipelagos (Preston 1962; MacArthur and Wilson 1967).

Arrhenius (1923a) also predicted that species-rich communities should have low Z values. Our results show that these arguments are dubious. The critical point is not that Z values are below or above 0.262 but that they are scale dependent. The C values of the power model also show an increasing trend over scales from small to large, so that they are also not scale independent (table 5). Our results support Connor and McCoy (1979, p. 815), who "are skeptical that any biological significance can be attached to these parameters and recommend that they be viewed simply as fitted constants devoid of specific biological meanings."

A Choice among Species-Area Models

Goodall (1952, p. 217) states that a decision to select a model "cannot be made on a priori grounds, but must rest on observational data." Connor and McCoy (1979) also suggest that the procedure to select a best model should be by fitting several models to real data and comparing results using statistical criteria. By expanding model (4) and going through the exercise of regression modeling, we have shown that the exponential model is only appropriate for small sampling areas, the power model is the best for intermediate sampling areas, and the logistic is the best for large-scale sampling. This result suggests that there objectively exits a "best" model for any particular data set and that there is no model that is universally best, all depending on sampling scales.

The exponential model was once widely accepted by ecologists, but it is largely ignored nowadays. The power model is the most widely used in the current literature, and researchers have presented it as a paradigm without questioning its appropriateness (Connor and McCoy 1979). The reason for the wide acceptance of the power model may lie in the fact that most field sampling programs are not large enough for the logistic curve to impose itself, and not so small as to make it clearly inadequate. For intermediate sampling areas, species-area data are frequently found to be satisfactorily fitted by the power model. Our results show, however, that it is necessary to question this paradigm before using it to describe data, except where there is sound ecological evidence supporting the power model as the appropriate form.

If sampling covers the whole of a community, the logistic is expected to be the best model to describe the species-area relationship. Contrary to the exponential model, which predicts a negative number of species when the area is very small, or the power model, which predicts an astonishingly large number of species when the area tends to be large, the logistic model predicts realistic numbers of species converging toward zero for an area vanishingly small and to some maximum for the whole community. Unfortunately, the logistic model has not received due attention in ecology and conservation biology, although it was proposed almost a half-century ago (Archibald 1949) and observed by various workers to fit large-scale sampling data (Vestal 1949; Kilburn 1963; Niering 1963; Whitehead and Jones 1969; Abbott 1973; Lassen 1975).

For large sampling areas, if only the exponential and the power models are compared, the former can frequently produce a better fit to observations than the latter, as shown by the Pasoh forest (table 2) and the Cole78 data (table 4). That

may be the reason that the exponential was once preferred by plant ecologists. However, if the logistic model is added to the comparison, it proves far superior to both the exponential and the power models for large sampling areas.

The random placement model (Coleman 1981) also describes the slow-rapid-slow accumulation of new species in a region; in other words, it is of the logistic type. Although that model is better than the exponential and power models, the logistic model (eq. [3]) gives a better fit in general. Remember, however, that species are rarely randomly distributed in nature. The nonrandomness of species distributions significantly affects species-area curves, which are related to the species' spatial distributions (Palmer and White 1994; He et al. 1996). There are two shortcomings to the random placement model: it is not useful for prediction because the total observed number of species is the maximum number of species predicted by the model, and construction of the model requires data on the abundance of all species in a region. We suggest that the random placement model should simply be considered as a null model under which species are randomly distributed over a region.

The Monod equation, which is also used to describe species-area relations (Clench 1979; Lauga and Joachim 1987; Colwell and Coddington 1994), is another model displaying an inflection point (in S–ln A space), just as the logistic and the random placement models do. The Monod equation is actually not a new form of species-area model but a special case of the logistic model obtained by imposing that $Z = 1$ in equation (3).

Estimating the Total Number of Species

One of the uses of species-area curves is to estimate the maximum number of species in a community or a type of vegetation through extrapolation (i.e., the number of species when $A \to \infty$). When doing so, two types of error are committed for the sake of generalization. The first one is statistical: a prediction is sought out of the range for which data are available. The second one is ecological: one assumes that the "real" community ideally occupies an area of infinite size where the conditions remain the same and where evolution has not given rise to other species. Although it corresponds to an essentialistic view of nature, this estimation by extrapolation seems of great interest to ecologists (Palmer 1990; Colwell and Coddington 1994). Many methods are available for this purpose, including curve fitting, relative abundance distributions, estimators derived from sampling theories, and nonparametric techniques (Bunge and Fitzpatrick 1993; Colwell and Coddington 1994). Our discussion here is confined to model fitting.

Behind all the methods to estimate the maximum number of species, an implied assumption is that in a limited area the number of species is finite. A community can hardly be completely sampled, but a community can be considered as "nearly completely sampled" if the sampling area tends to be large while the environmental conditions remain the same or else if the community is repeatedly sampled on several occasions. Neither the exponential nor the power model can appropriately be used to estimate the maximum number of species by extrapolation in a region,

because in these models S does not tend to some maximum when A increases. In contrast, the logistic model provides an upper-bound number of species when the area tends to be large. This property is of significance in conservation biology. When a community is thoroughly sampled or $A \rightarrow \infty$, the logistic model estimates the maximum number of species in that community to be B/C in equation (3). For example, from the fitted logistic model, the total number of tree species in the Pasoh forest can be estimated to be $4.438/0.00534 = 831$. This estimate is identical to that obtained by assuming a lognormal distribution of species abundances (F. He, unpublished data), and it exceeds the observed number of species (825) by only six species. This means that if the sampling area in the Pasoh forest was expanded to be much larger than 50 ha while remaining in the same type of landscape and habitat—but not far enough for evolution to produce more species—or if all the tree species in the Pasoh forest were thoroughly and correctly identified, the number of species would likely increase to be close to 831. Although the observed number of free-standing tree species in the Pasoh forest is the result of a supposedly complete and accurate census, it remains possible that some species were missed because of their rarity or misidentification.

From the logistic model, we predicted that the largest islands in the Pymatuning Lake (Coleman et al. 1982) can accommodate as many as 39 residing birds based on the 1978 data (the observed number on the largest island was 34) or 37 species based on the 1979 data (the number recorded on the largest island was 35).

As a caveat, we have to point out that the precision of the predictions by any model, either parametric or nonparametric, depends on the completeness of the observations, sampling scales, and taxonomic groups. For example, it is a mistake to use a data set about a particular group of species to attempt to predict the number of all species of a more encompassing group in a region. When observations are fitted by the logistic (or some other) model, one implicitly assumes that the community properties, species composition, and environment heterogeneity are defined for these particular data, and only for these data. Prediction of the number of species by extrapolation is done under the condition that all these properties remain the same as defined in the original data while area expands to be large. This process is not practically true, but conceptually reasonable.

LITERATURE CITED

Abbott, I. 1973. Birds of Bass Strait. Proceedings of the Royal Society of Victoria 85:197–223.

———. 1974. Numbers of plant, insect and land bird species on 19 remote islands in the Southern Hemisphere. Biological Journal of the Linnean Society 6:143–152.

Archibald, E. E. A. 1949. The specific character of plant communities. II. A quantitative approach. Journal of Ecology 37:260–274.

Arrhenius, O. 1921. Species and area. Journal of Ecology 9:95–99.

———. 1923a. Statistical investigations in the constitution of plant associations. Ecology 4:68–73.

———. 1923b. On the relation between species and area: a reply. Ecology 4:90–91.

Baltanás, A. 1992. On the use of some methods for the estimation of species richness. Oikos 65: 484–492.

Barkman, J. J. 1989. A critical evaluation of minimum area concepts. Vegetatio 85:89–104.

Boecklen, W. J., and N. J. Gotelli. 1984. Island biogeographic theory and conservation practice: species-area or specious-area relationships? Biological Conservation 29:63–80.

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

Cain, S. A., and G. M. de O. Castro. 1959. Manual of vegetation analysis. Harper, New York.

Clench, H. 1979. How to make regional lists of butterflies: some thoughts. Journal of the Lepidopterists' Society 33:216–231.

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

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

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

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

Connor, E. F., E. D. McCoy, and B. J. Cosby. 1983. Model discrimination and expected slope values in species-area studies. American Naturalist 122:789–796.

de Candolle, A. 1855. Géographie botanique raisonnée: ou l'exposition des faits principaux et des lois concernant la distribution géographique des plates de l'époque actuelle. Maisson, Paris.

Dorrer, E. 1976. F distribution: algorithm 322, collected algorithms from ACM. Vol. 2. Association for Computing Machinery, New York.

Dutilleul, P., and P. Legendre. 1993. Spatial heterogeneity against heteroscedasticity: an ecological paradigm versus a statistical concept. Oikos 66:152–171.

Engen, S. 1977. Exponential and logarithmic species-area curves. American Naturalist 111:591–594.

Evans, F. C., P. J. Clark, and R. H. Brand. 1955. Estimation of the number of species present on a given area. Ecology 36:342–343.

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.

Gleason, H. A. 1922. On the relation between species and area. Ecology 3:158–162.

———. 1925. Species and area. Ecology 6:66–74.

Goodall, D. W. 1952. Quantitative aspects of plant distribution. Biological Review 27:194–245.

Grassle, J., and N. Maciolek. 1992. Deep-sea species richness: regional and local estimates from quantitative bottom samples. American Naturalist 139:313–341.

Gunnill, F. C. 1982. Effects of plant size and distribution on the number of invertebrate species and individuals inhabiting the brown alga Pelvetia fastigiata. Marine Biology 69:263–280.

He, F., P. Legendre, C. Bellehumeur, and J. V. LaFrankie. 1994. Diversity pattern and spatial scale: a study of a tropical rain forest of Malaysia. Environmental and Ecological Statistics 1: 265–286.

He, F., P. Legendre, and J. V. LaFrankie. 1996. Spatial pattern of diversity in a tropical rain forest in Malaysia. Journal of Biogeography 23:57–74.

———. In press. Pattern of tree species in a tropical rain forest of Malaysia. Journal of Vegetation Science.

Hopkins, B. 1955. The species-area relations of plant communities. Journal of Ecology 43:409–426.

———. 1957. The concept of minimal area. Journal of Ecology 45:441–449.

Hubbell, S. P., and R. B. Foster. 1983. Diversity of canopy trees in a Neotropical forest and implications to conservation. Pages 25–41 in S. L. Sutton, T. C. Whitmore, and A. C. Chadwick, eds. Tropical rain forest: ecology and management. Blackwell Scientific, Oxford.

Jaccard, P. 1902. Lois de distribution florale dans la zone alpine. Bulletin de la Société Vaudoise des Sciences Naturelles 38:69–130.

———. 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des Sciences Naturelles 44:223–270.

Jobson, J. D. 1991. Applied multivariate data analysis. Vol. 1. Springer, New York.

Kilburn, P. D. 1963. Exponential values for the species-area relation. Science (Washington, D.C.) 141:1276.

———. 1966. Analysis of species-area relation. Ecology 47:831–843.

Kochummen, K. M., J. V. LaFrankie, and N. Manokaran. 1991. Floristic composition of Pasoh Forest Reserve, a lowland rain forest in Peninsular Malaysia. Journal of Tropical Forest Science 3:1–13.

Lassen, H. H. 1975. The diversity of freshwater snails in view of the equilibrium theory of island biogeography. Oecologia (Berlin) 19:1–8.

Lauga, J., and J. Joachim. 1987. L'échantillonnage des populations d'oiseaux par la méthode des E. F. P.: intérêt d'une étude mathématique de la courbe de richesse cumulée. Acta Oecologica/Oecologia Generalis 8:117–124.

Lawrey, J. D. 1991. The species-area curve as an index of disturbance in saxicolous lichen communities. Bryologist 94:377–382.

Loehle, C. 1990. Proper statistical treatment of species-area data. Oikos 57:143–145.

MacArthur, R. H., and E. O. Wilson. 1963. An equilibrium theory of insular zoogeography. Evolution 17:373–387.

———. 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.

May, R. M. 1975. Patterns of species abundance and diversity. Pages 81–120 in M. L. Cody and J. M. Diamond, eds. Ecology and evolution of communities. Belknap, Cambridge, Mass.

McGuinness, K. A. 1984. Equations and explanations in the study of species-area curves. Biological Review 59:423–440.

Niering, W. A. 1963. Terrestrial ecology of Kapingamarangi Atoll, Caroline Islands. Ecological Monographs 33:131–160.

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

Palmer, M. W., and P. S. White. 1994. Scale dependence and the species-area relationship. American Naturalist 144:717–740.

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

———. 1962. The canonical distribution of commonness and rarity. Ecology 43:185–215, 410–432.

SAS. 1989. SAS/STAT user's guide. Vol. 2. SAS Institute, Cary, N.C.

Schoener, T. W. 1976. The species-area relation within archipelagos: models and evidence from island land birds. Pages 629–642 in H. J. Frith and J. H. Calaby, eds. Proceedings of the 16th International Ornithological Conference. Australian Academy of Science, Canberra.

Simberloff, D. S., and E. O. Wilson. 1969. Experimental zoogeography of islands: the colonization of empty islands. Ecology 50:278–296.

Soulé, M. E., B. A. Wilcox, and C. Holtby. 1979. Benign neglect: a model of faunal collapse in the game reserves of East Africa. Biological Conservation 15:259–272.

Sugihara, G. 1981. $S = CA^Z, Z \approx 1/4$: a reply to Connor and McCoy. American Naturalist 117:790–793.

Vestal, A. G. 1949. Minimum areas for different vegetations: their determination from species-area curves. Illinois Biological Monograph 20:1–129.

Werff, H., van der. 1983. Species number, area and habitat diversity in the Galapagos Islands. Vegetatio 54:167–175.

Whitehead, D. R., and C. E. Jones. 1969. Small islands and the equilibrium theory of insular biogeography. Evolution 23:171–179.

Williams, C. B. 1943. Area and number of species. Nature (London) 152:264–267.

———. 1964. Patterns in the balance of nature. Academic Press, New York.

Williamson, M. 1981. Island populations. Oxford University Press, Oxford.

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

FANGLIANG HE*

CANADIAN FOREST SERVICE
   PACIFIC AND YUKON REGION
      506 W. BURNSIDE ROAD
         VICTORIA, BRITISH COLUMBIA V8Z 1M5
            CANADA

DEPARTMENT OF MATHEMATICS AND STATISTICS
   UNIVERSITY OF VICTORIA
      VICTORIA, BRITISH COLUMBIA V8W 3P4
         CANADA

PIERRE LEGENDRE†

DÉPARTEMENT DE SCIENCES BIOLOGIQUES
   UNIVERSITÉ DE MONTRÉAL
      CP 6128, SUCC. CENTRE-VILLE
         MONTRÉAL, QUÉBEC H3C 3J7
            CANADA

* E-mail: fhe@a1.pfc.forestry.ca.
† E-mail: Legendre@ere.umontreal.ca.