Strong and weak tests of macroecological theory

Brian McGill, Dept of Ecology and Evolutionary Biology, Univ. of Arizona, Tucson, AZ 85721, USA (mail@brianmcgill.org).

Macroecology is a rapidly growing branch of ecology. The essence of macroecology can be summarized as a two-step process: 1) find large-scale patterns and 2) find the explanations/mechanisms for those patterns. Much of the work on step 1 has focused on identifying the shape of various curves. This includes some of the most famous of all macroecological patterns: the power-law ($S = cA^z$) species area relationship (SPAR), the hollow curve distribution of species abundances (SAD), and the skewed-lognormal distribution of body size (BSD). A large number of papers have debated the correct curve shapes. For example it has been suggested that SPARs should really be logarithmic, sigmoidal, or asymptotic (such as the Michaelis–Menton) instead of the power law (Connor and McCoy 1979, Sugihara 1981, Connor et al. 1983, Williams 1995, Lomolino 2000, 2002, Lomolino and Weiser 2001, Williamson et al. 2001). Over two-dozen different distributions have been suggested for the SAD (Pielou 1977, Tokeshi 1993). Moreover, much of the work on step 2 has proceeded by testing theories according to whether they produce curves of the correct shape (i.e. the shape observed in nature). In short, it is fair to say that macroecology has been practiced as a study of curve shapes.

I argue that this focus on curve shapes is unfortunate and even counterproductive for macroecology. To see why, let us explore in more detail how macroecological mechanisms are identified. Unfortunately, macroecological questions are, almost by definition, too large to perform the type of replicated, manipulative, controlled experiments that form the backbone of scientific progress in much of the rest of ecology. Thus, the search for mechanism usually proceeds through theory. A theory is built which suggests that if certain mechanisms are important, then certain predictions should be true. If the predictions fail, the mechanisms lose credence; if they succeed, the mechanisms accumulate support. Of course, the strength of this method depends on the quality of the predictions. In practice, often the sole

prediction made is the shape of the curve (with parameters free to maximize fit).

The weakness of this mode of testing is most painfully obvious in the search for mechanisms in one of the oldest and most well known patterns in macroecology: the species abundance distribution (SAD). The SAD describes the relative abundances of different species within a community – many rare species and a few highly abundant species (known as a hollow curve). Well over two dozen different curves have been suggested as the "right" curve with more coming every year (Pielou 1977, Tokeshi 1993, Sichel 1997, Harte et al. 1999, Dewdney 2000, Hubbell 2001). This work on SADs has proceeded as I previously described: a hypothesized mechanism leads to a theory that makes one prediction about curve shape. Macroecologists ought to worry that dozens of theories have been proposed and all fit the data. Macroecologists must also worry that few if any curves (and their mechanisms) have been decisively rejected. In short, this mode of testing has failed to reject most of the dozens of proposed theories.

In the rest of this paper, I will first try to demonstrate that testing exclusively by curve fitting is very weak and a poor way for macroecology to proceed in the search for mechanisms. Then, I will look at some ways to make stronger tests.

Why curve fitting is a weak test

There are at least four reasons why testing by curve fitting is weak.

Reason no. 1 – multiple explanations

More than one mechanism can produce any given set of data. Pielou noted this in the context of SADs (1977, p. 123). She pointed out that for many probability distributions, mathematical theorems show there are multi-

ple mechanistic hypotheses that can lead to a given distribution. In a similar spirit, Cohen (1968) derived two biological models of SADs which produce a broken stick distribution (MacArthur 1957). Cohen produces broken-stick distributions despite starting from biological assumptions very different from MacArthur's. One of these models starts from assumptions almost the opposite of MacArthur's, yet ends up producing the same distribution.

Reason no. 2 – free parameters

When parameters are left free and then set to maximize the fit to the data, they surely ought to produce a good fit most of the time. This is vividly captured in a famous quote that I have seen attributed to various sources including Linus Pauling and Euler which says, "give me three free parameters and I will draw you an elephant, give me four and I will wag the trunk."

The ability of curve fitting with free parameters to fit curves that are fundamentally not a match can be seen in the following example, which I draw from SPARs. We know that a SPAR must be a monotonically increasing, decelerating function. I will now show through a simple Monte Carlo simulation that the power law ($S = cA^z$) will fit the data of any functional form that meets the basic requirements of being a monotonically increasing, decelerating function regardless of all other attributes. I generated data from five different underlying functions: power law ($S = cA^z$), logarithm ($S = \log(c + zA)$), two with an asymptote (Michaelis–Menton and Ivlev) and one sigmoidal (logistic). All five of these functions are monotonic increasing, decelerating and have two free parameters. I then added noise to the data. Finally, I tried fitting two different functional forms (power law, Michaelis–Menton) to this noisy data, and calculated a mean-corrected non-linear r^2.

The results are presented in Table 1 and 2. Overall, we can see that any functional form will fit any other functional form quite well ($r^2 > 0.89$ and usually $r^2 > 0.98$). The goodness of the fit depends heavily on where

Table 1. Fit of the power and Michaelis–Menton functions to various underlying models with noise. Average of nonlinear, mean-centered r^2 across 100 simulations presented in each cell. The amount of noise and number of points sampled had very small effects ($\Delta r^2 < 0.012$). We can see that either function fits any other function "well". It is still possible to make distinctions between grossly different curves (e.g. fits to data based on logistic fit worse than others) and MM fits the similar Ivlev function better than the power, as we would expect given the similarity in functional forms.

	Power	MM	Log	Logit	Ivlev
MM	0.990	1.000	1.000	0.890	0.995
Power	1.000	0.970	0.996	0.932	0.947

Table 2. Fits of the power function depend on the range explored. The XRange column indicates spatial extent. The logit functions had an inflection point at 1, while the MM function reached its half-saturation level at x = 1. Thus if the scale ran from 0.5 to 1.5 (i.e. XRange = 1) centered around the intermediate region or even from 0 to 2 (i.e. XRange = 2), then relatively little of the asymptotic region was sampled. At this scale, all functions fit well. When the x-scale ran from 0 to 10, then the region from 2 to 10 (80% of the sample) was very close to the asymptote. At this scale, functions with an asymptote fit the power function markedly worse than functions without an asymptote (Power, Log)

XRange	Power	MM	Log	Logit	Ivlev
1	1	0.989	0.999	0.989	0.975
2	1	0.97	0.995	0.938	0.947
10	1	0.878	0.988	0.895	0.648

we sample along the x-axis (Table 2). Not surprisingly, the fit is better when we don't sample the region where the asymptotes occur. The fit is exceedingly good if we don't sample this region, and still decent even if we do sample it moderately.

Thus, through the power of curve fitting, functional forms which are fundamentally different from the known underlying functional form still fit the data very well. The only place this breaks down is if we have an asymptote and we sample a very large region of the asymptote (e.g. 80% of the sample range is in the asymptote). Curve fitting with free parameters offers no hope for differentiating among subtly different theories.

Reason no. 3 – there is no such thing as a curve that fits "best"

To illustrate this point, I will give some data from a brief empirical example. I used data from the North American Breeding Bird Survey or BBS (Robbins et al. 1986, Price et al. 1995) averaged over 5 years to eliminate noise. I randomly sampled 100 routes that meet certain quality criteria. I fit five distributions commonly used as a SAD using either moment matching or maximum likelihood methods (Evans et al. 1993). I then compared the goodness of fit using 9 different measures: Kolmogorov-Smirnov, likelihood, three based on different measures comparing the actual and observed CDF using various nonlinear versions of r^2 (Wilkinson 1997), and four based on a χ^2 test using different binning schemes. All of the measures of fit that I used have been previously applied to this problem with the possible exception of likelihood. Some measures emphasize fit over the whole range of the data, others emphasize fit in a region of special interest (e.g. where most of the probability density is). Depending on the context, any of these differences might be desirable.

The results are presented in Table 3. Usually, the lognormal performed the best, but the "best distribution" depended on the measure of fit that we use. The

Table 3. Number of times each distribution was the best fit for various measures. The "best" distribution is highlighted in bold for each measure of goodness-of-fit. The 1st χ^2 is based on 10 bins on a arithmetic scale. The 2nd is based on Preston-type binning on a log scale. The 3rd used 10 bins on an arithmetic scale for the lowest 80% and one bin for the rest. The 4th column used 5 bins on an arithmetic scale. Notice that the "best" fitting distribution depends on the measure of goodness of fit used.

	R2	R2MC	R2Corr	K	L	χ^2	Log χ^2	χ^2 bin+1	χ^2 5 bins
MLE power	0	0	0	0	0	2	0	0	0
Mom power	1	1	2	0	0	1	0	0	1
Lognormal	28	28	**28**	62	100	32	6	27	33
Trunc LN	**34**	**34**	17	37	0	49	27	67	47
LogSeries	15	15	27	0	0	15	47	4	13
Neg binomial	18	18	17	1	0	1	11	1	5

lognormal, truncated lognormal, and logseries are all the "best distribution" depending on the measure of fit. There is no a priori reason why we should choose one measure over another. This suggests it will be hard to come up with rigorous scientific tests based on the concept of "best fit to the data."

Reason no. 4 – statistical limit theorems

Most biologists know that the central limit theorem (CLT) of statistics (Grimmett and Stirzaker 1992) states that we should find normal distributions any time many variables interact additively. In addition, most macroecologists are familiar with the fact that by simple properties of the logarithm, the CLT predicts that when many variables interact multiplicatively, we should find a lognormal distribution (McAlister 1879, MacArthur 1960, May 1975). Two common processes in ecology produce interaction of multiplicative random variables. One is a "sequence-of-hurdles". Fitness is a good example (see also Shockley 1957): $W = p_1 p_2 p_3 p_4...$ where p_1 is the probability of surviving to maturity, p_2 is the probability of surviving the first winter, p_3 is the probability of holding a territory, p_4 is the probability of finding a mate, etc. Thus, $\log W = \Sigma \log p_i$ should be normally distributed. The other common process is "compound-growth" (e.g. Caswell 2001). In compound growth, the final state, say population size N, is given by $N_t = N_0 \lambda_1 \lambda_2 \lambda_3...\lambda_t$. Hence, the final state is the product of independent random variables (i.e. the λ_i's which here represent annual growth rates of a population but might also be growth rates of an individual's body size).

A distribution that is closely related to the lognormal is the power or Pareto or Zipf-Mandelbrot distribution (Mandelbrot 1982, Schroeder 1991, Evans et al. 1993). The power distribution can either be derived as a limit of the lognormal or through various statistical arguments in its own right. The lognormal distribution converges to the power distribution as the variance goes to infinity. This can be shown analytically (West and Shlesinger 1989) and graphically (Fig. 1). A formula calculating the degree of error (West and Shlesinger 1989) shows that they are extremely close over many

orders of magnitude. We expect the variance to be high in ecological systems both because the number of variables involved is high (which by the CLT leads to high variance) and because ecological systems are notoriously noisy. There are also deep CLT-like statistical reasons for why we might find a power distribution (Simon 1955, Mandelbrot 1963, Bak et al. 1987, West and Shlesinger 1989, Dewdney 2000, Allen et al. 2001, Bianconi and Barabasi 2001). In practice, because the power distribution is the limit of the lognormal and because they are so close to each other even when variance is fairly small, it is often difficult to distinguish between them empirically. And theoretically it is unnecessary to do so. Hence, I will refer to these distributions collectively as POLO (power lognormal) throughout the remainder of this paper.

Why do these POLO limit theorem arguments mean that curve fitting is a weak test? First, recall our context of testing mechanisms by producing theories containing mechanisms which match the data. The CLT POLO arguments show that:

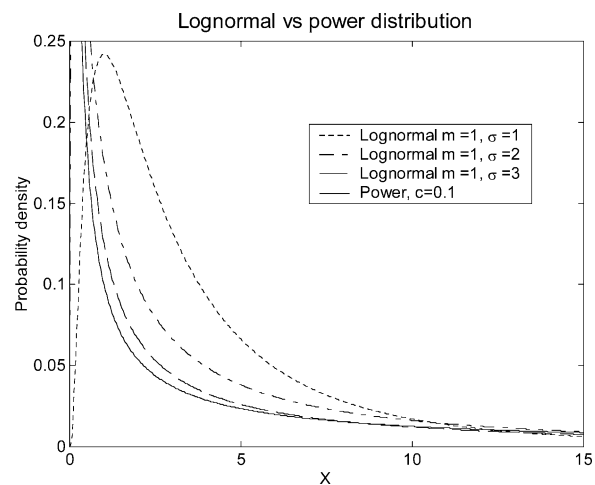1. *Data will be POLO distributed.* Most of the classical patterns in macroecology are close to POLO distri-

Fig. 1. The probability density function (PDF) for the power distribution and low- to high-variance lognormal distributions. Note how the high variance lognormal distribution is similar in shape to the power distribution.

butions – e.g. lognormal distributions of abundance (Preston 1948, Whittaker 1965), body size (Maurer and Brown 1988, Brown and Nicoletto 1991), and size of geographic range (Pagel et al. 1991). Even the deviations from lognormal such as the "excess" of rare species are often universal (e.g. found in distribution of human incomes or the sizes of cities) and come from CLT-like arguments (Montroll and Shlesinger 1982).

2. *Any complex theory that invokes multiplication of complex factors will produce POLO-like distributions.* I like to think of this as the "central limit theorem of theories". Although I know of no exact mathematical theorem stating this, arguments such as the central limit theorem suggest that theories which are complex enough and invoke mechanisms similar enough to those of the central limit theorem (e.g. multiplication of independent random variables) will converge in some limit to a CLT distribution (such as the lognormal). Thus, Sugihara's random fraction model of SADs (1980) converges to the lognormal (Pielou 1975) as do many others (e.g. Engen and Lande 1996). Similarly, a great many different theories have found a power distribution as the limit. The broken stick distribution (MacArthur 1957) converges to the power distribution (Perline 1996), as does the logseries distribution when $x \to 1$.

When we combine points 1 and 2 above together, we see that almost any theory will match almost any data as long as we only look at the shape of the distribution; that shape will usually be POLO in nature! So, it is impossible to test a mechanistic theory merely by producing a theory which generates the shape of the data if there is a CLT argument for that shape.

Making curve fitting a stronger test

None of the deficiencies of the previous section can be repaired – we cannot simply improve curve fitting to address them. The best solution is to develop new methods of testing. I do this in the next section. Yet, even within the domain of curve fitting, we can make our results much stronger.

I suggest that there is a hierarchy of progressively stronger definitions of what it means to fit "well":

Level 1) The curve fits the data well from an "eyeball" point of view (i.e. human perception).

Level 2) The curve fits the data well according to *some objective measure* such as r^2 or χ^2.

Level 3) The curve fits the data *better than the appropriate null hypothesis* by some objective measure.

Level 4) The curve is *statistically significantly better*

than the appropriate null hypothesis (i.e. we can reject the null).

Level 5) The curve is statistically significantly better than the appropriate null hypothesis *after penalization for extra parameters*.

In most areas of ecology, level 4 is the standard required to publish a paper. In macroecology, levels 1 and 2 have often sufficed. For example, Hubbell's recent book (2001) works primarily at level 1. Dewdney (2000) differs by actually going to level 3.

Because many fundamentally different curves can have an objectively good fit, we can begin to detect the "right" curve only in comparison with other curves. This suggests that level 3 is the lowest acceptable level. But because there are a large number of measures and the "best" fit can depend on the measure, we should probably require more than just a preponderance of evidence (level 3) and should require level 4 instead.

Level 3 and above require a null hypothesis. We have already seen that a commonly appropriate null hypothesis is the POLO. Other arguments suggests that the Poisson (the law of rare events, Taylor and Karlin 1998) and the power law (which fits a wide variety of data as demonstrated in point 2) are also common null hypotheses.

Simply achieving level 3 would be much stronger than most of the results reported to date in macroecology (and I expect eliminate a great many of the proposed distributions for SADs). However, the usual standard for publishing in ecology is level 4. There are many possible tests of whether a curve fits significantly better than a null hypothesis. Some of these include:

- F statistics – the ratio of sums of squares (e.g. Hilborn and Mangel 1997)
- Likelihood ratios – the ratio of the probabilistic likelihood under two different models is asymptotically distributed as a χ^2 distribution (e.g. Hilborn and Mangel 1997).
- Monte Carlo and repeated samples – if we generate many repetitions either through random simulation methods or from many datasets (e.g. different locales), we can see if the proposed model fits better than the null hypothesis 95% of the time.
- Kolmogorov–Smirnov statistic – when using the Kolmogoro–Smirnov it is usually important to apply a Lilliefors type correction (1967), since the parameters are usually estimated from the data.
- χ^2 test – these tests involve a loss of power due to the need to bin continuous data, and there is no a priori preferred "best" binning method.

The first two methods can only be used on nested models. The last two can be used only to reject the possibility that we have sampled from a null-hypothesis distribution. It cannot accept an alternative to the null. The Akiake and Bayesian information criteria address

these problems and also take us up to level 5 by using likelihood ratios and applying a penalty factor for using more parameters (Hilborn and Mangel 1997).

In short, tests based on curve fitting can be made reasonably rigorous if we use existing statistical machinery. It is a mystery why most macroecologists have chosen to ignore this well-known machinery for so long.

Moving beyond curve fitting – strong tests in macroecology

A heuristic to help us achieve stronger tests is to return to two of the identified problems with curve fitting: the central limit theorem and the malleability of free fitting parameters. If we can address these issues, we will have stronger tests. The following four methods of testing theories have no free parameters and are not predicted by the CLT:

1) Constancy or other patterns in the parameters
2) A priori prediction of the parameters
3) The temporal and spatial scale of exchangeability
4) Correlations and other side predictions

The CLT, by making no predictions in these four areas, affords us four additional methods of testing a macroecological theory. Each of these four methods improves the quality of the test beyond mere curve-fitting. Even at stage one in macroecology (pattern identification), it may be desirable to explore these rather than the shapes of curves. At stage two (tests of mechanisms), their use is essential to provide stronger deductive tests. I will now give examples of each of these four tests.

Example of 1). For a long time, it was thought that all SPARs had z-values approximately equal to 0.25 (May 1975). This would have been a very interesting and informative fact – at least as important as the power-law shape of the curves. But Connor and McCoy (1979) noticed that the z-values have a much larger range, from about 0.05 to about 1.0. Rosenzweig examined these z-values and noticed that in fact z-values were strongly correlated with the temporal scale of the processes involved in creating the SPAR. Intercontinental SPARs have a $z \approx 1.0$; SPARs between islands have a $z \approx 0.25$; and nested SPARS on the mainland have $z \approx 0.15$. By looking at patterns in the parameters, Rosenzweig was able to start sorting out different scales and mechanisms behind the SPARs.

Example of 2). As far back as 1962, Preston developed a theory that attempted to predict the parameters of a curve. He suggested that SADs follow a very specific form of the lognormal curve which he called the canonical distribution. From this he deduced that the power coefficient (z-value) of the SPAR should be about 0.26. As the number of SPARs with slopes strongly different than 0.26 has increased, this theory has come to be considered disproved (Rosenzweig 1995) and some faulty assumptions have been identified (Leitner and Rosenzweig 1997). But this exactly proves the power of the approach. Had Preston merely shown that his canonical lognormal distribution produced a power-law SPAR (i.e. a curve shape with free fitting parameters), his theory would have been unfalsifiable.

Example of 3). The idea of exchangeability comes from statistical mechanics and probability. Two entities are considered exchangeable if they are identical from the point of view of the question. Looking at the spatial and temporal scales over which species are and are not exchangeable can be quite useful. For example, the abundances in species abundance distributions (SADs) have been found to often remain constant over large spatial and temporal scales (McGowan and Walker 1985, Lawton and Gaston 1989, Ebeling et al. 1990, Boucot 1996, Brett et al. 1996, Hadly and Maurer 2001). This suggests that any mechanism purporting to explain SADs must operate on similarly large spatial and temporal scales. For example, there is nothing in most of the classical niche preemption models (Motomura 1932) to suggest that the same species would preempt first at each different site and hence attain high abundance everywhere, yet we find that species that are abundant at one location are often abundant elsewhere. Of course, we could start imagining that arrival and preemption occur on a very large spatial scale, but this is a good example of the utility of looking at spatial and temporal scales. Similarly, the core-satellite hypothesis (Hanski 1982) while not explicit about temporal scales, presumably treats species as exchangeable over fairly short time scales – a problem for systems where species abundance is constant over very long time scales.

Example of 4). It is a truism that any theory that makes multiple predictions can be tested more strongly. Rosenzweig and Abramsky (1997) points out that even the weakest prediction (a qualitative yes or no) can be part of a robust test if we have enough of them – he calls this a "dipswitch test". Hence, a good theory in macroecology will make multiple predictions, many of which are unrelated to the shape of curves. For example, various correlations or the precise value of certain measurements might be predicted as well. As a specific example, a theory of SADs would have to be considered very strong if it could predict not just the shape of the curve, but which species would be common or rare based on individual species traits.

Although this paper has primarily focused on the SAD, it should be noted that its conclusions and implications apply to any field of macroecology that is focusing on the shape of curves. For example, as already mentioned, considerable energy is currently going into debating the correct shape of the species area relationship (SPAR). Our example of the indistin-

guishability of curves suggested that we may never resolve this empirically. Even on a theoretical level, fractals and colored noise give a CLT-like argument for the existence of power laws (Mandelbrot 1982, Tokeshi 1999). Ultimately, pursuing the four areas mentioned above may better advance our understanding of SPARs than debating the shape of curves. For example, Rosenzweig (1995) has made considerable progress by combining techniques one and three and examining different temporal scales over which parameters show a regular pattern.

Macroecology is a vitally important field of ecology today. It promises much hope in unraveling fundamental principles as well as being a useful tool for conservationists. It would be a shame if this utility were to be considered unreliable because we macroecologists do not use rigorous tests of our theories.

References

Allen, A. P., Li, B.-L. and Charnov, E. L. 2001. Population fluctuations, power laws and mixtures of lognormal distributions. – Ecol. Lett. 4: 1–3.

Bak, P., Tang, C. and Wiesenfeld, K. 1987. Self-organized criticality: an explanation of 1/f noise. – Phys. Rev. Lett. 59: 381–384.

Bianconi, G. and Barabasi, A. L. 2001. Competition and multiscaling in evolving networks. – Europhys. Lett. 54: 436–442.

Boucot, A. J. 1996. Epilogue. – Palaeogeogr. Palaeoclimatol. Palaeoecol. 127: 339–359.

Brett, C. E., Ivany, L. C. and Schopf, K. M. 1996. Coordinated stasis: an overview. – Palaeogeogr. Palaeoclimatol. Palaeoecol. 127: 1–20.

Brown, J. H. and Nicoletto, P. F. 1991. Spatial scaling of species composition: body masses of North American land mammals. – Am. Nat. 138: 1478–1512.

Caswell, H. 2001. Matrix population models: construction, analysis, and interpretation. – Sinauer Associates.

Cohen, J. E. 1968. Alternate derivations of a species-abundance relation. – Am. Nat. 102: 165.

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

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

Dewdney, A. K. 2000. A dynamical model of communities and a new species–abundance distribution. – Biol. Bull. 198: 152–165.

Ebeling, A. W., Holbrook, S. J. and Schmitt, R. J. 1990. Temporally concordant structure of a fish assemblage: bound or determined? – Am. Nat. 135: 63–73.

Engen, S. and Lande, R. 1996. Population dynamic models generating the lognormal species abundance distribution. – Math. Biosci. 132: 169–183.

Evans, M., Hastings, N. and Peacock, B. 1993. Statistical distributions. – John Wiley & Sons.

Grimmett, G. R. and Stirzaker, D. R. 1992. Probability and random processes. – Clarendon Press.

Hadly, E. A. and Maurer, B. A. 2001. Spatial and temporal patterns of species diversity in montane mammal communities of western North America. – Evol. Ecol. Res. 3: 477–486.

Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. – Oikos 38: 210–221.

Harte, J., Kinzig, A. P. and Green, J. 1999. Self-similarity in the distribution and abundance of species. – Science 284: 334–336.

Hilborn, R. and Mangel, M. 1997. The ecological detective: confronting models with data. – Princeton Univ. Press.

Hubbell, S. P. 2001. A unified theory of biodiversity and biogeography. – Princeton Univ. Press.

Lawton, J. H. and Gaston, K. J. 1989. Temporal patterns in the herbivorous insects of bracken: a test of community predictability. – J. Anim. Ecol. 58: 1021–1034.

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

Lilliefors, H. W. 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. – J. Am. Stat. Assoc. 64: 399–402.

Lomolino, M. V. 2000. Ecology's most general, yet protean pattern: the species area relationship. – J. Biogeogr. 27: 17–26.

Lomolino, M. V. 2002. Comment on M. Williamson, K. J. Gaston and W. M. Lonsdale. – J. Biogeogr. 29: 555–557.

Lomolino, M. V. and Weiser, M. D. 2001. Towards a more general species–area relationship: diversity on all islands, great and small. – J. Biogeogr. 28: 431–445.

MacArthur, R. 1957. On the relative abundance of bird species. – Proc. Natl Acad. Sci. USA 43: 293–295.

MacArthur, R. 1960. On the relative abundance of species. – Am. Nat. 94: 25–36.

Mandelbrot, B. B. 1963. New methods in statistical economics. – J. Polit. Econ. 71: 421–440.

Mandelbrot, B. B. 1982. The fractal geometry of nature. – W. H. Freeman and Co.

Maurer, B. A. and Brown, J. H. 1988. Distribution of energy use and biomass among species of North American terrestrial birds. – Ecology 69: 1923–1932.

May, R. M. 1975. Patterns of species abundance and diversity. – In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Belknap Press of Harvard Univ. Press, pp. 81–120.

McAlister, D. 1879. The law of the geometric mean. – Proc. R. Soc. Lond. 29: 367–376.

McGowan, J. A. and Walker, P. W. 1985. Dominance and diversity maintenance in an oceanic ecosystem. – Ecol. Monogr. 55: 103–118.

Montroll, E. and Shlesinger, M. F. 1982. On 1/f noise and other distributions with long tails. – Proc. Natl Acad. Sci. USA 79: 3380–3383.

Motomura, L. 1932. On the statistical treatment of communities. – Zool. Mag. 44: 379–383.

Pagel, M. D., May, R. M. and Collie, A. R. 1991. Ecological aspects of the geographical distribution and diversity of mammalian species. – Am. Nat. 137: 791–815.

Perline, R. 1996. Zipf's law, the central limit theorem and the random division of the unit interval. – Phys. Rev. E 54: 220–223.

Pielou, E. C. 1975. Ecological diversity. – John Wiley & Sons.

Pielou, E. C. 1977. Mathematical ecology. – John Wiley & Sons.

Preston, F. W. 1948. The commonness and rarity of species. – Ecology 29: 254–283.

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

Price, J., Droege, S. and Price, A. 1995. The summer atlas of North American birds. – Academic Press.

Robbins, C. S., Bystrak, D. and Geissler, P. H. 1986. The breeding bird survey: its first fifteen years, 1965–1979. – US Dept of the Interior, Fish and Wildlife Service.

Rosenzweig, M. L. 1995. Species diversity in space and time. – Cambridge Univ. Press.

Rosenzweig, M. L. and Abramsky, Z. 1997. Two gerbils of the Negev: a long-term investigation of optimal habitat selection and its consequences. – Evol. Ecol. 11: 733–756.

Schroeder, M. 1991. Fractals, chaos, and power laws: minutes from an infinite paradise? – W. H. Freeman and Co.

Shockley, W. 1957. On the statistics of individual variations of productivity in research laboratories. – Proc. Inst. Radio Engineers 45: 279–290.

Sichel, H. S. 1997. Modelling species-abundance frequencies and species-individual functions with the generalized inverse Gaussian–Poisson distribution. – S. Afr. Stat. J. 31: 13–37.

Simon, H. A. 1955. On a class of skew distribution functions. – Biometrika 42: 425–440.

Sugihara, G. 1980. Minimal community structure: an explanation of species-abundance patterns. – Am. Nat. 116: 770–787.

Sugihara, G. 1981. $S = CAz$, $z = 1/4$: a reply to Connor and McCoy. – Am. Nat. 117: 790–793.

Taylor, H. M. and Karlin, S. 1998. An introduction to stochastic modelling. – Academic Press.

Tokeshi, M. 1993. Species abundance patterns and community structure. – Adv. Ecol. Res. 24: 111–186.

Tokeshi, M. 1999. Species coexistience. – Blackwell Sciences Ltd.

West, B. J. and Shlesinger, M. F. 1989. On the unbiquity of 1/f noise. – Int. J. Modern Physics B 3: 795–819.

Whittaker, R. H. 1965. Dominance and diversity in land plant communities. – Science 147: 250–260.

Wilkinson, L. 1997. SYSTAT 7.0 Statistics. – SPSS.

Williams, M. R. 1995. An extreme value function model of the species incidence and species-area relations. – Ecology 76: 2607–2616.

Williamson, M., Gaston, K. J. and Lonsdale, W. M. 2001. The species-area relationship does not have an asymptote. – J. Biogeogr. 28: 827–830.