

Species Abundance in a Forest Community in South China: A Case of Poisson Lognormal Distribution

Zuo-Yun YIN^{1,2}, Hai REN^{1*}, Qian-Mei ZHANG¹, Shao-Lin PENG^{1,3}, Qin-Feng GUO^{1,4} and Guo-Yi ZHOU¹

(1. South China Botanical Garden (formerly South China Institute of Botany), the Chinese Academy of Sciences, Guangzhou 510650, China;

2. Guangdong Forest Research Institute, Guangzhou 510520, China;

3. School of Life Sciences, Zhongshan University, Guangzhou 510275, China;

4. US Geological Survey, Jamestown, ND 58401, USA)

Abstract: Case studies on Poisson lognormal distribution of species abundance have been rare, especially in forest communities. We propose a numerical method to fit the Poisson lognormal to the species abundance data at an evergreen mixed forest in the Dinghushan Biosphere Reserve, South China. Plants in the tree, shrub and herb layers in 25 quadrats of 20 m×20 m, 5 m×5 m, and 1 m×1 m were surveyed. Results indicated that: (i) for each layer, the observed species abundance with a similarly small median, mode, and a variance larger than the mean was reverse J-shaped and followed well the zero-truncated Poisson lognormal; (ii) the coefficient of variation, skewness and kurtosis of abundance, and two Poisson lognormal parameters (σ and μ) for shrub layer were closer to those for the herb layer than those for the tree layer; and (iii) from the tree to the shrub to the herb layer, the σ and the coefficient of variation decreased, whereas diversity increased. We suggest that: (i) the species abundance distributions in the three layers reflects the overall community characteristics; (ii) the Poisson lognormal can describe the species abundance distribution in diverse communities with a few abundant species but many rare species; and (iii) $1/\sigma$ should be an alternative measure of diversity.

Key words: Dinghushan Biosphere Reserve; forest community; maximum likelihood estimate (MLE); reverse J-shaped curve; species abundance distribution (SAD); value-trying method; zero-truncated Poisson lognormal (PLN) distribution.

Species abundance (SA) relationship is a fundamental focus in community ecology, once referred to as the “minimal community structure” (Sugihara 1980; Molles 1999). Research in this area has advanced greatly in recent years (for a review, see Peng *et al.* 2003). The species abundance distribution (SAD) is commonly presented in the form of a histogram with S_r , the number of species containing r members, as the ordinate and r as the abscissa. Presented in this manner, S_r is thus the frequency of frequency (Bulmer 1974; Slocomb

et al. 1977; Pielou 1985). In general, a community or biological assemblage has a few common species, but many rare species (quite often singletons), with the former greatly exceeding the latter in total number of individuals (Fisher *et al.* 1943; Preston 1948, 1962a, 1962b; Whittaker 1965; May 1975; Krebs 1978; Pielou 1985; Brown 1995; McGill 2003a, 2003b; McGill and Collins 2003; Chave 2004). The “distribution of commonness and rarity” among species described first by Preston (1948, 1962a, 1962b) has been regarded as

Received 24 Aug. 2004 Accepted 21 Dec. 2004

Supported by the National Natural Science Foundation of China (30270282, 30200035 and 021627), the Forestry Science and Technology Research Planning of Guangdong Province of China (2003-2), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-SW-120).

* Author for correspondence. Tel (Fax): +86 (0)20 3725 2916; E-mail: <renhai@scib.ac.cn>.

one of the best-documented patterns in natural communities (Molles 1999). The distribution is truncated, with a missing zero class, because S_0 , the number of species not represented at all in the collection, is usually unknown (Bulmer 1974; Pielou 1985).

Early in the 1940s, statistical models were proposed to describe the SAD (Fisher *et al.* 1943; Preston 1948) and, today, these models still generate considerable interest (Magurran and Henderson 2003; McGill 2003a, 2003b, 2003c; Chave 2004). In an attempt to find a mathematical expression to describe the SADs, Preston (1948) first introduced a left-truncated lognormal (LN) distribution to fit the SA data grouped on a log scale. He called this log-transformed abundance an "octave". Since then, the LN has been used widely to model a variety of SA datasets from various communities or collections, such as moths, birds, snakes, and plants, and has attracted increasing attention (Whittaker 1965; Gauch and Chase 1974; Slocumb *et al.* 1977; Coleman 1981; Miller and Wiegert 1989; Xie *et al.* 1995; Engen and Lande 1996; Basset *et al.* 1998; Hill and Hamer 1998; Watt 1998; Harte *et al.* 1999; Fesl 2002; McGill 2003a, 2003b, 2003c; Yin *et al.* 2005). Several approaches have been taken to fit the Gaussian (or LN on a log scale), including the variation of parameters algorithm (Gauch and Chase 1974), maximum likelihood estimate (MLE; Slocumb *et al.* 1977), and the general non-linear least squares method (Yin and Liao 1999; Yin *et al.* 2005).

Previous attempts to fit the LN distribution to the SA data (Preston 1948; Bliss 1965; Yin and Liao 1999) have treated the observed distribution as a truncated, grouped LN distribution. However, this statistical treatment ignored the effect of the Poisson sampling variability and adopted more-or-less arbitrary abundance grouping. Furthermore, this method does not provide a satisfactory approximation to the probabilities, P_r , when r is near to zero, as was shown by Grundy (1951) and reiterated by Cassie (1962) and Kempton and Taylor (1974). To overcome these drawbacks, some have proposed a mathematically more accurate compound distribution, the Poisson lognormal (PLN; Grundy 1951;

Cassie 1962; Pielou 1969, 1985; Bulmer 1974; Kempton and Taylor 1974; Etienne and Olf 2004). In fitting the PLN (here zero truncated), Bulmer (1974) derived the MLE of its parameters and developed a computer program in Algol. Kempton and Taylor (1974) developed another MLE. In their fitting procedure, the PLN probability, P_r , can be obtained for $r < 32$ by numerical integration using Simpson's rule; for higher values of r , the continuous approximation was used. Comparison of their results with those of Bliss (1965) indicated that Bliss's use of the continuous approximation for r close to zero gave inexact parameter estimates.

However, practically, these methods are somewhat difficult to use and, thus the application of this distribution has been limited. Moreover, case studies on the PLN distributions of SAs have been rare relative to the widely used LN, especially in forest communities (Peng *et al.* 2003). In the present study, we use the zero-truncated PLN to fit the SA data surveyed from a lower subtropical forest in South China. We develop a numerical method to obtain the MLE of the parameters and examine the ecological implications of these parameters.

1 Materials and Methods

1.1 Data source

We investigated a 1-hm² (100 m×100 m) permanent forest site using the methods of the Smithsonian/MAB (Man and Biosphere programme) biodiversity program (Dallmeier 1992) at an elevation of approximately 250 m in the Dinghushan Biosphere Reserve (112°30'39"–112°33'41" E, 23°09'21"–23°11'30" N), Guangdong, South China, in 1999 (Zhang *et al.* 2002). The site was established in 1992. The vegetation was a lower subtropical evergreen needle- and broad-leaved mixed forest dominated by *Castanopsis chinensis* Hance, *Schima superba* Gardn. et Champ., and *Pinus massoniana* Lamb. We surveyed the tree species (diameter at breast height, i.e. DBH ≥ 1 cm) in the tree layer at 25 adjacent quadrats of 20 m×20 m. Each such quadrat was divided, in turn, into 16 smaller quadrats of 5 m×5 m, within which we then randomly chose

one quadrat in which to record the shrubs and seedlings of trees (DBH <1 cm, height ≥ 50 cm) in the shrub layer. Meanwhile, we randomly chose a quadrat of 1 m \times 1 m within each of the 25 (5 m \times 5 m) quadrats and counted the herbs and woody seedlings (height <50 cm) in the herb layer.

1.2 Poisson lognormal distribution

We denote the observed number of species in the collection by S and the expected total number of species that may have been represented by S^* ; the latter quantity is usually unknown and must be estimated from the data. Suppose that: (i) the number of members of the j th species in the collection is a Poisson variate with a mean, λ_j ($j = 1, 2, \dots, S^*$); and (ii) the values of λ_j can be regarded as constituting S^* independent observations from an LN distribution (i.e. $\ln \lambda_j$ is normally distributed with a mean μ and variance σ^2). Thus, the probability that a randomly chosen species represented by r individuals in the collection is a discrete, compound PLN distribution with the probability function (Grundy 1951; Pielou 1969; Bulmer 1974):

$$P_r = \int_0^{\infty} \left(\frac{\lambda^r e^{-\lambda}}{r!} \right) \left\{ \frac{1}{\lambda \sigma \sqrt{2\pi}} \exp \left[-\frac{(\ln \lambda - \mu)^2}{2\sigma^2} \right] \right\} d\lambda \quad (1)$$

where $r = 0, 1, 2, 3, \dots$. We then obtain the expected number of species with r individuals in the collection, $S_r = S^* P_r$.

Normally, the value of S^* is unknown because some very rare species may be missing in an observation, but it can be estimated from the SA data. Because $S_0 = S^* P_0$ and $S^* - S = S_0$, we can obtain the estimate of S^* containing those rare species not yet collected in the sample (Bulmer 1974; Pielou 1985):

$$S^* = S/(1-P_0) \quad (2)$$

where

$$P_0 = \int_0^{\infty} \frac{1}{\lambda \sigma \sqrt{2\pi}} \exp \left[-\lambda - \frac{(\ln \lambda - \mu)^2}{2\sigma^2} \right] d\lambda \quad (3)$$

which is the probability of having no individuals. Thus, the PLN distribution without such a category has the probability function:

$$P_r' = P_r/(1-P_0) \quad (r = 1, 2, 3, \dots) \quad (4)$$

which is called the zero-truncated PLN with two parameters, σ and μ .

1.3 Data analysis

Based on the per species count data collected at the site, we denoted the observed SAD as S_1, S_2, \dots, S_r , which were the numbers of species represented by 1, 2, \dots , and r individuals, respectively. To fit the truncated PLN to the SA data, we proposed a numerical method termed the "value-trying method" to estimate the parameters σ and μ through the MLE. We chose appropriate ranges of the two parameters and searched for their optimum estimates within these ranges by designing an iterative computer program in Mathematica 4 (Wolfram 1999), such that the log likelihood function was maximized (Bulmer 1974):

$$\ln L(\sigma, \mu) = \sum_{r=1} S_r \ln P_r' = \max \quad (5)$$

The iteration by equation 5 started from initial values of two parameters that were estimated by fitting a continuous LN distribution to the data on a log scale (Preston 1948; Bliss 1965; Slocomb *et al.* 1977) or by using the moment-matching method. This iteration proceeded with each parameter's step length being 0.001, and terminated when the maximum value of $\ln L(\sigma, \mu)$ reached default accuracy in Mathematica 4 (i.e. absolute error $< 10^{-13}$ in all cases).

Because the data are discrete, we can apply the goodness-of-fit Chi-squared (χ^2) test to compare the observed frequency distribution with the expected one (e.g. Bulmer 1974; Kempton and Taylor 1974; Glover and Mitchell 2001; McGill 2003a). We conducted the χ^2 test with $c-3$ degrees of freedom (d.f.) where c is the number of categories used for the χ^2 statistic calculation (e.g. Bulmer 1974). As usual, the frequencies with small expectations are pooled, so that no expectation is less than 5 and, in situations with 1 d.f., χ^2 should have a 0.5 correction for discontinuity (Glover and Mitchell 2001). Because the test is always right tailed, if the P value associated with the χ^2 test, $P(\chi^2)$, is greater than the significance level 0.05, or the calculated χ^2 statistic is under the critical value for the test performed at this level with $c-3$ d.f., then the null

hypothesis H_0 (that the observed frequency distribution is consistent with the expected) is accepted (Glover and Mitchell 2001).

2 Results

2.1 Survey data and descriptive statistics

The survey results of the present study indicated that: (i) the tree layer had a total of 3 890 individuals from 69 species in 25 quadrats (20 m×20 m), with the main dominant tree species including *Castanopsis chinensis* Hance, *Schima superba* Gardn. et Champ., and *Pinus massoniana* Lamb.; (ii) the shrub layer had 428 individuals from 39 species in 25 quadrats (5 m×5 m), with the most dominant shrubs being *Litsea rotundifolia* var. *oblongifolia* (Nees) Allen, *Psychotria rubra* (Lour.) Poir., and *Rhodomyrtus tomentosa* (Ait.) Hassk.; and (iii) the herb layer contained 151 individuals from 32 species in 25 quadrats (1 m×1 m), with the most dominant herb being *Lophatherum gracile* Brongn. (Table 1). The community was an evergreen needle- and broad-leaved mixed secondary forest typical of the lower subtropical forest vegetation in the monsoon climate of South China, particularly in the proximity of the Tropic of Cancer.

The number of individuals (i.e. abundance, r) per

species had a similar median ($= 2-3$) and the same mode ($= 1$) for each of the three layers in the community sampled. The variance/mean ratio (V/M) of the variable r for each layer had large values, with the highest value in the tree layer and the lowest in the herb layer. Statistics without measurement units (i.e. coefficient of variation (CV), skewness, and kurtosis) displayed a similar pattern: the tree layer had the largest values, whereas the shrub and herb layers had similar values, although the herb layer had the lowest CV and skewness and the shrub layer had the lowest kurtosis. Yet, the difference in CV and skewness among the three layers was relatively small.

2.2 Species diversity and rank abundance curves

The forest community studied showed high plant species diversity and evenness measured by Shannon-Wiener's and Simpson's indices, which increased from the tree layer to the shrub layer to the herb layer, although the differences were relatively small (Table 2). This was also indicated by the rank abundance curves, which showed a similar steepness for the three layers, with the steepest slope for the herb layer (Fig. 1).

2.3 Species abundance distribution

The SADs among the three layers showed some apparent similarities. Plotting the number of species,

Table 1 Species abundance data sampled in the tree, shrub, and herb layers of the evergreen needle- and broad-leaved mixed forest in the Dinghushan Biosphere Reserve, South China

Layer	Area (m ²)	S	N	Min	Max	Med	Mode	Mean	SD	V/M	CV	Sk	Ku
Tree	10 000	69	3 890	1	772	3	1	56.38	138.99	342.65	2.47	3.60	13.94
Shrub	625	39	428	1	83	3	1	10.97	18.47	31.07	1.68	2.66	6.97
Herb	25	32	151	1	29	2	1	4.72	5.94	7.47	1.26	2.60	8.40

CV , coefficient of variation; Ku, kurtosis; Min, Max, Med, minimum, maximum, and median of abundance, respectively; SD , standard deviation; Sk, skewness; S , N , observed total numbers of species and individuals, respectively; V/M , variance/mean ratio.

Table 2 Diversity and evenness indices and estimates of the zero-truncated Poisson lognormal distributions fitted to the species abundance in the three layers of the evergreen mixed forest in the Dinghushan Biosphere Reserve, South China

Layer	D_{SW}	E_{SW}	D_{Sim}	E_{Sim}	σ	μ	Chi-squared test			P_0	S^*
							χ^2	d.f.	$P(\chi^2)$		
Tree	3.923	0.642	0.899	0.912	3.868	-2.385	3.767	5	0.583	0.673	211
Shrub	4.010	0.759	0.904	0.927	2.290	-0.386	3.703	2	0.157	0.478	75
Herb	4.207	0.841	0.921	0.951	1.728	-0.413	2.660	1	0.103	0.483	62

D_{SW} , D_{Sim} , diversity measured by Shannon-Wiener's and Simpson's indices, respectively (D_{SW} is calculated using \log_2); d.f., degrees of freedom; E_{SW} , E_{Sim} , evenness measured by Shannon-Wiener's and Simpson's indices, respectively (E_{SW} is calculated using \log_2); σ , μ , two distribution parameters; P_0 , probability of no individuals; S^* , the theoretical total number of species.

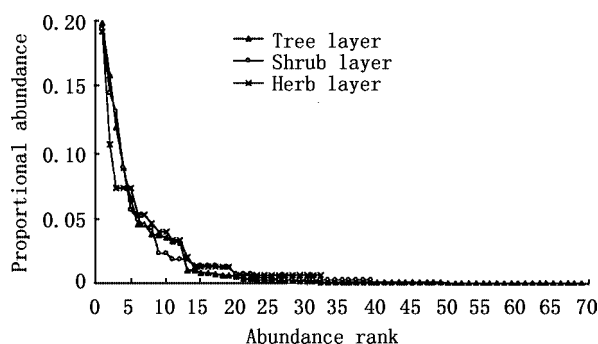


Fig. 1. Rank abundance curves of the tree, shrub, and herb layers of the evergreen needle- and broad-leaved mixed forest in the Dinghushan Biosphere Reserve, South China.

S_r , with r individuals against r , we obtained the reverse J-shaped frequency distributions of observed SAs in all three layers. Because the variance of r was much larger than the mean for each layer (Table 1), we attempted to use the PLN distribution, which could accommodate this case, to describe the observed SAD. Examining these SA data in the three layers revealed that they all fitted well the zero-truncated PLN distributions because all P values for χ^2 tests, $P(\chi^2)$, were much greater than 0.05 (Table 2). This was also shown in the comparisons between the observed and expected frequency distributions on a log scale equivalent to Preston's octave scale (Fig. 2).

Nonetheless, there were some distinctions between the three layers. The two parameters σ and μ for the shrub layer were relatively close to those for the herb layer, but somewhat different from those for the tree layer, as were P_0 , the probability with no individuals, and S^* , the expected total number of species in the universe (Table 2). The σ value for the tree layer was greater than that for the shrub and herb layers, with the lowest value obtained for the herb layer. The values of CV and skewness showed similar trends. The $|\mu|$, as well as the kurtosis, was in the order: tree layer > herb layer > shrub layer, with the last two close to each other (Tables 1, 2). The steepness of the reverse J-shaped probability histograms of the fitted

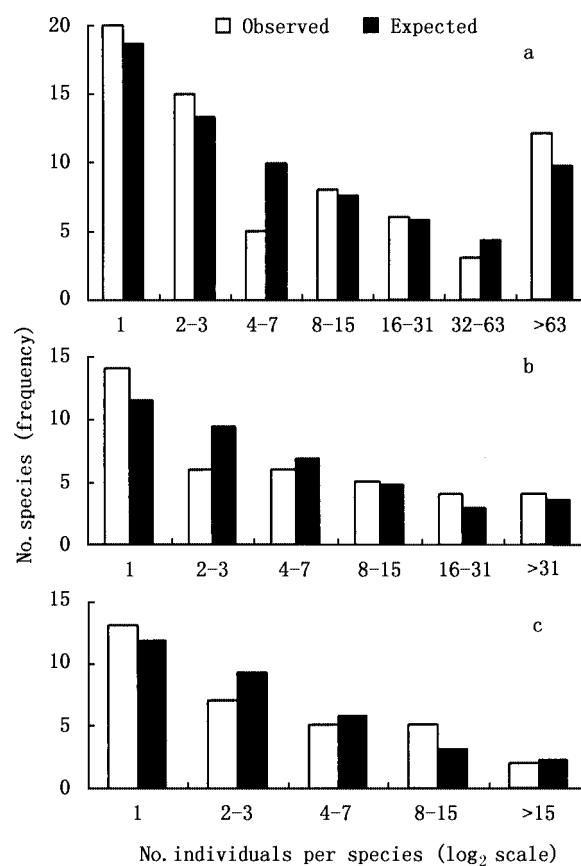


Fig. 2. Species abundance distributions on a log scale (for grouping method, see Kempton and Taylor 1974; Krebs 1978) in the (a) tree, (b) shrub, and (c) herb layers of the evergreen mixed forest in the Dinghushan Biosphere Reserve, South China. Open bars indicate the observed number of species; filled bars indicate the number of species expected from the fitted zero-truncated Poisson lognormal distribution.

zero-truncated PLN displayed the following order: tree layer < shrub layer < herb layer (Fig. 3). That is, the tree layer had the lowest expected proportion of rare species and the herb layer had the highest, with the shrub layer in the middle. However, the abundant species showed a reverse pattern. This was similar to the case for the observed proportion of species: in the tree, shrub, and herb layers, species with one individual accounted for 29.0%, 35.9%, and 40.6%, respectively, whereas species with more than 16 individuals accounted for 29.0%, 20.5%, and 3.1%, respectively.

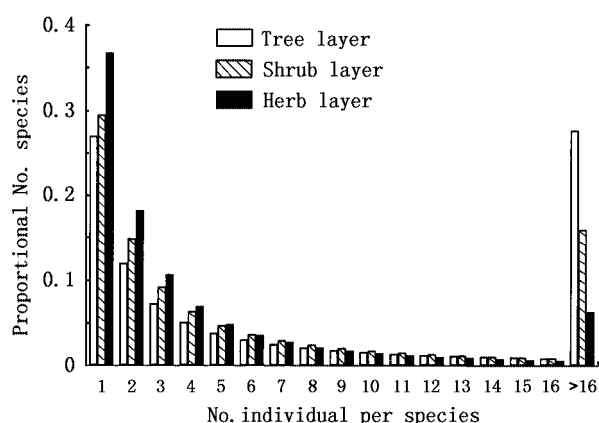


Fig. 3. Comparison of the expected proportional numbers of species from the fitted zero-truncated Poisson log-normal distributions among the three layers of the evergreen mixed forest in the Dinghushan Biosphere Reserve, South China.

3 Discussion

The SA relationships among the three layers of the forest community had some similarities. The SAD for either the tree, shrub, or herb layers displayed a compound zero-truncated PLN distribution with the variance much greater than the mean. This is because the community had high species diversity (Peng and Wang 1995; Peng 1996) with several codominant species, rather than low diversity with a mono-dominant species. In the former case, the LN (and, thus, PLN, as discussed below) provides a good fit, whereas in the latter case, a log-series distribution is more appropriate (for a review, see Peng *et al.* 2003). For example, in the tree layer, five dominant species each had >5% of the total of 3 890 individuals, three species had >10%, and two species had >15%; in the shrub layer, six species had >5% of the total of 428 individuals and three species had >10%; and in the herb layer, seven species had >5% of the total of 151 individuals and two species had >10%. This suggests that each layer of the community had several codominant species and many rare species. In this case, the PLN can provide a good fit, as can the LN.

Moreover, the abundance per species in the three layers had similar, small median ($= 2-3$) and mode

($= 1$) values and had similar CV and skewness. The steepness of the rank abundance curve (and, thus, species diversity) was also similar and the SAD showed similar, reverse J-shapes for each layer. These similar properties of the SA relationships among the three layers indicates some similarity of the structure and integrity as components of a whole community. This may be due to the fact that all the species in the different layers in the same community had been inhabited in and, thus, had adjusted to the same macrohabitat (i.e. the lower subtropical monsoon climate, during a long-term succession).

However, there were some differences in the SA relationships among the three layers. The SAD in the shrub layer was more-or-less close to the one in the herb layer, but relatively different from the one in the tree layer, as reflected by the values of CV, skewness, kurtosis, and the two fitted PLN parameters σ and μ (and, thus, P_0 and S^* ; Tables 1,2). This could be because the species in both the shrub and herb layers were located in the microhabitat at a somewhat different location to the tree layer: the former were found in the understory of the forest (relatively shady and wet), whereas the latter was found in the overstory (relatively sunny and dry). Yet, these differences may also be related to the different total sampling areas (10 000, 625, and 25 m² for the tree, shrub, and herb layers, respectively). Studies have shown that the total number of species increases with sampling area, as indicated by the species-area relationship (Arrhenius 1921; Hopkins 1955; Preston 1962a, 1962b; Kilburn 1966; MacArthur and Wilson 1967; Connor and McCoy 1979; Sugihara 1980; Pielou 1985; Williamson 1988; Brown 1995). Consequently, with increasing sampling area, the proportion of rare species (such as singletons) may decrease, whereas that of common species may increase, as shown in Fig. 3. However, different situations may exist: with increasing sampling effort, the proportion of rare species also increases, whereas that of common species decreases, or one or two of these proportions remain constant. In the case that the proportions of both rare and common species remain

constant with different (but adequately intensive) sampling efforts, the differences in the SADs between the understory and overstory should result from microhabitat differences. This could only be resolved by sampling the same areas for all the layers, although our sampling areas are adequately large for the three layers (Dallmeier 1992).

The PLN parameter σ is the population parameter, but μ is dependent on the sampling fraction and, hence, the size of the sampling unit (Kempton and Taylor 1974). Simulations using Mathematica-4 programs indicate that the two parameters of the PLN distribution together determine its location and shape but, overall, μ is its location parameter, whereas σ is its shape parameter. The distribution curve can be either reverse J-shaped or single hump-shaped (unimodal). Particularly, when $\mu \leq 0$, for all $\sigma > 0$, the slope of the curve at any point is negative. This is the reason why the SADs for the three layers are all reverse J-shaped (Fig. 3). Furthermore, if $\mu \leq 0$, the smaller the σ value and the steeper the reverse J-shaped curve. This may explain why the fitted SAD became steeper with decreasing σ from the tree, to the shrub, to the herb layers (Table 2; Fig. 3).

A steeper SAD curve suggests that there is a larger number of rare species (especially singletons) and fewer common species in a collection. This may mean a flatter rank abundance curve and, thus, a comparatively greater equitability or evenness of allotment of individuals among species, leading to an increase in species diversity. This can be demonstrated by the present case study: like the evenness or diversity index based on Shannon-Wiener's or Simpson's index, the steepness of the curve increased from the tree layer to the shrub layer and to the herb layer. Shannon-Wiener's and Simpson's indices are two of the most popular measures of diversity (Krebs 1978; Mi and Ma 2003). The former is adopted from information theory, whereas the latter is derived from probability theory (Krebs 1978). In our case, the σ value shows a negative correlation with both indices; that is, the smaller the σ , the steeper the curve and, therefore, the higher the diversity (Table

2; Fig. 3). Furthermore, the σ value decreased with the CV from the tree to the shrub to the herb layer (Tables 1, 2). A smaller CV (a statistic measuring the degree of mean variation of a random variable; e.g. the abundance r in the present study) implies a greater evenness of abundance distribution among species and, thus, an increase in species diversity.

A diversity measure combines the total number of species and their relative abundances within a community. Kempton and Taylor (1974) suggested that the PLN parameters S^* and σ , independent of the sample size, reflect these two aspects of the diversity, respectively. Kempton and Taylor (1974) illustrated that S^* and σ separately have an evidently very poor power to discriminate diversity between sites, whereas the two function of the two parameters, namely $D = \log S^* + C \log \sigma$, is a very powerful diversity discriminant. They further obtained $C = -1.69$ by maximizing the between- to within-sites mean square ratio for D . From another viewpoint, Bulmer (1974) derived a reasonable estimate of Shannon-Wiener's information diversity index based on the assumption of the PLN (i.e. $H = \log S^* - \sigma^2/2$). These ecologists all showed that diversity was positively correlated with S^* but was negatively correlated with σ . However, this is not completely the case in the present study. From the tree, to the shrub, to the herb layer, regardless of the decrease in S^* (in particular, S^* for the tree layer was approximately threefold S^* for the other two layers), the diversity increased slightly as σ decreased (Table 2). This suggests that, at least in our case, the increasing diversity may result from a decreasing σ , not from an increasing S^* . For these reasons, we suggest that $1/\sigma$ be used as an alternative measure of diversity. The larger the value of $1/\sigma$, the more diverse the community. Such a suggestion would require further confirmation using additional SA datasets.

With respect to the estimation of PLN parameters, in addition to the MLE used in the present study, several other methods, such as moment estimate and the least squares method, can also be adopted. Using the method of moment matching, we obtain the most convenient but roughest estimates of the two parameters

that can be used as the initial values of the other two methods to obtain more precise estimates. Using a continuous approximation of the LN distribution to the log-scaled SA data (Preston 1948; Bliss 1965), we can also obtain initial values of the parameters σ and μ to be used in the iteration of fitting the PLN. Interestingly, our estimated values of σ and μ (Table 2) almost fall into the ranges of the same PLN parameters estimated by Kempton and Taylor (1974) in 72 different moth catches for 18 sites for 4 yr: σ is 0.99 to 2.92 and μ is -4.38 to 2.16.

Because the LN distribution of the SA presented by Preston (1948) is an approximation to the PLN (Grundy 1951; Bliss 1965; Pielou 1985) and can often describe any diverse community with a large number of species that perform diverse ecological functions (May 1975; Peng *et al.* 2003), the PLN should do the same. Krebs (1978, 2001) summarized that the LN distribution fits a variety of data from species-rich communities. Based on the central limit theorem of statistics, the LN of the SAD is expected to arise in diverse communities when many independent factors interact multiplicatively (MacArthur 1960; May 1975; McGill 2003c). The LN can easily be explained through the biological concept of niche partitioning (Bulmer 1974; May 1975; Pielou 1975; Sugihara 1980). Engen and Lande (1996) derived the SAD of an LN type from population dynamic models based on the intra- or inter-specific density regulation and correlated environmental noise. Because PLN is a more appropriate mathematical expression of the LN for describing the SAD, the possible mechanisms for the LN of the SAD should suit well with PLN.

Over the past two decades, Hubbell (1979, 1997, 2001) and others have developed a unified neutral theory to explain patterns of biodiversity and biogeography, such as the SAD, in community ecology. This theory is neutral because it assumes that individuals of all species within a community are functionally equivalent. This is in contrast with classical adaptive, niche-based theories that stress functional differences between species. The other central assumption of Hubbell's theory is that the community is saturated with

individuals and the SAD should follow a zero-sum multinomial (ZSM) distribution. Within Hubbell's approximation, the distribution of an LN-like shape is a particular case of the ZSM when the immigration coupling between the metacommunity and the local community is lower. The neutral theory is considered by many ecologists as a radical shift from established niche theories, but has been hotly debated (e.g. McGill 2003a; Volkov *et al.* 2003). McGill (2003a) showed that, at least for some data, neutral theory performs worse than the traditional, simpler null model (i.e. the LN). However, this is not the end for neutral theory; rather, it is the "end of the beginning" (Nee and Stone 2003). Chave (2004) suggests that the niche and neutral theories should be complementary, not conflicting, and the future challenge may be to integrate them. Regardless of its origins, the LN (and, thus, PLN) is important because it allows us to predict the SAD within a community and it also provides other valuable insights into the community organization, such as species-area relationships and succession (Preston 1962a, 1962b; Bazzaz 1975; May 1975; Miller and Wiegert 1989; Aoki 1995; Molles 1999).

Acknowledgements The authors thank Zhong-Liang HUANG, Shi-Zhong LIU, Wan-Hui YE, Qing-Fa YU, and You-Chang ZHANG in South China Botanical Garden, the Chinese Academy of Sciences, and many others for field assistance.

References

- Aoki I (1995). Diversity and rank-abundance relationship concerning biotic compartments. *Ecol Model* **82**, 21–26.
- Arrhenius O (1921). Species and area. *J Ecol* **9**, 95–99.
- Basset Y, Novotny V, Miller S, Springate ND (1998). Assessing the impact of forest disturbance on tropical invertebrates: Some comments. *J Appl Ecol* **35**, 461–466.
- Bazzaz FA (1975). Plant species diversity in old-field successional ecosystems in southern Illinois. *Ecology* **56**, 485–488.
- Bliss CI (1965). An analysis of some insect trap records. In: Patil GP, ed. *Classical and Contagious Discrete Distributions*. Statistical Publishing Society, Calcutta. pp. 385–397.

- Brown JH (1995). *Macroecology*. University of Chicago Press, Chicago, IL.
- Bulmer MG (1974). On fitting the Poisson lognormal distribution to species abundance data. *Biometrics* **30**, 101–110.
- Cassie RM (1962). Frequency distribution models in the ecology of plankton and other organisms. *J Anim Ecol* **31**, 65–92.
- Chave J (2004). Neutral theory and community ecology. *Ecol Lett* **7**, 241–253.
- Coleman BD (1981). On random placement and species-area relations. *Math Biosci* **54**, 191–215.
- Connor EF, Mccoy ED (1979). The statistics and biology of the species–area relationship. *Am Nat* **113**, 791–833.
- Dallmeier F (1992). Long-term monitoring of biological diversity in tropical forest areas: Methods for establishment and inventory of permanent plots. In: UNESCO, ed. *MAB Digest* 11. UNESCO, Paris. pp. 72.
- Engen S, Lande R (1996). Population dynamic models generating the lognormal species abundance distribution. *Math Biosci* **132**, 169–183.
- Etienne RS, Olff H (2004). A novel genealogical approach to neutral biodiversity theory. *Ecol Lett* **7**, 170–175.
- Fesl C (2002). Biodiversity and resource use of larval chironomids in relation to environmental factors in a large river. *Freshwater Biol* **47**, 1065–1087.
- Fisher RA, Corbet AS, Williams CB (1943). The relation between the number of species and the number of individuals in a random sample from an animal population. *J Anim Ecol* **12**, 42–58.
- Gauch HG, Chase GB (1974). Fitting the Gaussian curve to ecological data. *Ecology* **55**, 1377–1381.
- Glover TJ, Mitchell KJ (2001). *An Introduction to Biostatistics*. McGraw-Hill, NY.
- Grundy PM (1951). The expected frequencies in a sample of an animal population in which the abundances of the species are log-normally distributed. Part I. *Biometrika* **38**, 427–434.
- Harte J, Kinzig A, Green J (1999). Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336.
- Hill JK, Hamer KC (1998). Using species abundance models as indicators of habitat disturbance in tropical forests. *J Appl Ecol* **35**, 458–460.
- Hopkins B (1955). The species area relations of plant communities. *J Ecol* **43**, 409–426.
- Hubbell SP (1979). Tree dispersion, abundance and diversity in a dry tropical forest. *Science* **203**, 1299–1309.
- Hubbell SP (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16** (Suppl.), 9–21.
- Hubbell SP (2001). *A Unified Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Kempton RA, Taylor LR (1974). Log-series and log-normal parameters as diversity discriminants for the Lepidoptera. *J Anim Ecol* **43**, 381–399.
- Kilburn PD (1966). Analysis of the species area relation. *Ecology* **47**, 831–843.
- Krebs CJ (1978). *Ecology: The Experimental Analysis of Distribution and Abundance*, 2nd edn. Harper & Row Publishers, NY.
- Krebs CJ (2001). *Ecology: The Experimental Analysis of Distribution and Abundance*, 5th edn. Benjamin Cummings, San Francisco, CA.
- MacArthur RH (1960). On the relative abundance of species. *Am Nat* **94**, 25–36.
- MacArthur RH, Wilson EO (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.
- Magurran AE, Henderson PA (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**, 714–716.
- May RM (1975). Patterns of species abundance and diversity. In: Cody MT, Diamond JM, eds. *Ecology and Evolution of Community*. Harvard University Press, Cambridge, MA. pp 81–120.
- McGill BJ (2003a). A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885.
- McGill BJ (2003b). Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecol Lett* **6**, 766–773.
- McGill BJ (2003c). Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685.
- McGill BJ, Collins C (2003). A unified theory for macroecology based on spatial patterns of abundance. *Evol Ecol Res* **5**, 469–492.
- Mi XC, Ma KP (2003). Advances in phytoecology in China II. Advances in plant synecology in China. *Acta Bot Sin* **45** (Suppl.), 70–76 (in Chinese with an English abstract).

- Miller RI, Wiegert RG (1989). Documenting completeness, species–area relations, and the species–abundance distribution of a regional flora. *Ecology* **70**, 16–22.
- Molles MC (1999). *Ecology: Concepts and Applications*. McGraw-Hill, NY.
- Nee S, Stone G (2003). The end of the beginning for neutral theory. *Trends Ecol Evol* **18**, 433–434.
- Peng SL (1996). *Community Dynamics in Lower Subtropical Forest*. Science Press, Beijing (in Chinese).
- Peng SL, Wang BS (1995). Forest succession at Dinghushan, Guangdong, China. *Acta Bot Sin* **7**, 75–80.
- Peng SL, Yin ZY, Ren H, Guo QF (2003). Advances in research on the species–abundance relationship models in multi-species collection. *Acta Ecol Sin* **23**, 1590–1605 (in Chinese with an English abstract).
- Pielou EC (1969). *An Introduction to Mathematical Ecology*. Wiley-Interscience, NY.
- Pielou EC (1975). *Ecological Diversity*. Wiley-Interscience, NY.
- Pielou EC (1985). *Mathematical Ecology*, 2nd edn. Wiley-Interscience, NY.
- Preston FW (1948). The commonness and rarity of species. *Ecology* **29**, 254–283.
- Preston FW (1962a). The canonical distribution of commonness and rarity. *Ecology* **43**, 185–215.
- Preston FW (1962b). The canonical distribution of commonness and rarity. *Ecology* **43**, 410–432.
- Slocumb J, Stauffer B, Dickson KL (1977). On fitting the truncated lognormal distribution to species–abundance data using maximum likelihood estimation. *Ecology* **58**, 693–696.
- Sugihara G (1980). Minimal community structure: An explanation of species abundance patterns. *Am Nat* **116**, 770–787.
- Volkov I, Banavar JR, Hubbell SP, Maritan A (2003). Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037.
- Watt AD (1998). Measuring disturbance in tropical forests: A critique of the use of species abundance models and indicator measures in general. *J Appl Ecol* **35**, 467–469.
- Whittaker RH (1965). Dominance and diversity in land plant communities. *Science* **147**, 250–260.
- Williamson M (1988). Relationship of species number to area, distance and other variables. In: Myers AA, Giller PS, eds. *Analytical Biogeography*. Chapman and Hall, London.
- Wolfram S (1999). *The Mathematica Book*, 4th edn. Cambridge University Press, Cambridge.
- Xie JY, Chen LZ, Ghirelli L, Chiesura LF (1995). Biodiversity studies on *Quercus ilex* woods in Veneto, Italy. *Acta Bot Sin* **37**, 386–393 (in Chinese with an English abstract).
- Yin ZY, Liao WB (1999). Studies on lognormal distribution patterns of species abundance of south subtropical forest community, China. *Guihaia* **19**, 221–224 (in Chinese with an English abstract).
- Yin ZY, Peng SL, Ren H, Guo Q, Chen ZH (2005). LogCauchy, log-sech and lognormal distributions of species abundances in forest communities. *Ecol Model* **148**, 329–340.
- Zhang QM, Huang ZL, Liu SZ, Zhou GY (2002). The community structure of coniferous and broad-leaved mixed forest in the Dinghushan. In: Zhou GY, Zhang QM, Wen DZ, Peng CL, Zhang DQ, eds. *Tropical and Subtropical Forest Ecosystem*, Vol. 9. China Environmental Science Press, Beijing. pp. 18–27 (in Chinese with an English abstract).

(Managing editor: Ya-Qin HAN)