

RESEARCH ARTICLE

Examining the patterns and dynamics of species abundance distributions in succession of forest communities by model selection

Zuo-Yun Yin^{1,2,3,4,5}*, Lu Zeng⁶, Shao-Ming Luo⁷*, Ping Chen¹, Xiao He⁸*, Wei Guo¹, Bailian Li⁵

1 College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong, China, **2** Guangdong Provincial Key Laboratory of Forest Breeding, Protection and Utilization, Guangzhou, Guangdong, China, **3** Guangdong Forest Research Institute (Guangdong Academy of Forestry), Guangzhou, Guangdong, China, **4** South China Botanical Garden of the Chinese Academy of Sciences, Guangzhou, Guangdong, China, **5** Ecological Complexity and Modeling Laboratory, Department of Botany and Plant Sciences, University of California, Riverside, California, United States of America, **6** Guangdong Ecological Engineering Vocational College, Guangzhou, Guangdong, China, **7** Faculty of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong, China, **8** College of Economics, Guangdong University of Finance and Economics (formerly Guangdong University of Business Studies), Guangzhou, Guangdong, China

* These authors contributed equally to this work.

* yinzuoyun@163.com (ZYY); smluo@gdut.edu.cn (SML); 807933187@qq.com (XH)



OPEN ACCESS

Citation: Yin Z-Y, Zeng L, Luo S-M, Chen P, He X, Guo W, et al. (2018) Examining the patterns and dynamics of species abundance distributions in succession of forest communities by model selection. PLoS ONE 13(5): e0196898. <https://doi.org/10.1371/journal.pone.0196898>

Editor: Tim A. Mousseau, University of South Carolina, UNITED STATES

Received: November 11, 2016

Accepted: April 23, 2018

Published: May 10, 2018

Copyright: © 2018 Yin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: This work was supported by the Education Department of Guangdong Province, China, Award Number: 2016KZDXM002 (Recipient: Zuo-Yun Yin, Ph.D.); Ministry of Science and Technology of the People's Republic of China, Award Number: 2012BAD22B05 (Recipient: Lu Zeng); Science and Technology Department of Guangdong Province, China, Award Number:

Abstract

There are a few common species and many rare species in a biological community or a multi-species collection in given space and time. This hollow distribution curve is called species abundance distribution (SAD). Few studies have examined the patterns and dynamics of SADs during the succession of forest communities by model selection. This study explored whether the communities in different successional stages followed different SAD models and whether there existed a best SAD model to reveal their intrinsic quantitative features of structure and dynamics in succession. The abundance (the number of individuals) of each vascular plant was surveyed by quadrat sampling method from the tree, shrub and herb layers in two typical communities (i.e., the evergreen needle- and broad-leaved mixed forest and the monsoon evergreen broad-leaved forest) in southern subtropical Dinghushan Biosphere Reserve, South China. The sites of two forest communities in different successional stages are both 1 ha in area. We collected seven widely representative SAD models with obviously different function forms and transformed them into the same octave (\log_2) scale. These models are simultaneously confronted with eight datasets from four layers of two communities, and their goodness-of-fits to the data were evaluated by the chi-squared test, the adjusted coefficient of determination and the information criteria. The results indicated that: (1) the logCauchy model followed all the datasets and was the best among seven models; (2) the fitness of each model to the data was not directly related to the successional stage of forest community; (3) according to the SAD curves predicted by the best model (i.e., the logCauchy), the proportion of rare species decreased but that of common ones increased in the upper layers with succession, while the reverse was true in the lower layers; and (4) the difference of the SADs increased between the upper and the

2015A040404045 (Recipient: Ping Chen); and Science and Technology Department of Guangdong Province, China, Award Number: 2016A020207003 (Recipient: Ping Chen).

Competing interests: The authors have declared that no competing interests exist.

lower layers with succession. We concluded that the logCauchy model had the widest applicability in describing the SADs, and could best mirror the SAD patterns and dynamics of communities and their different layers in the succession of forests. The logCauchy-modeled SADs can quantitatively guide the construction of ecological forests and the restoration of degraded vegetation.

Introduction

There are a few common species and many rare species in a biological community (or a multi-species collection) in given space and time. This hollow (i.e., inverse J-shaped, hyperbolic) distribution curve of the species number on the individual number (abundance) in the community is called “species abundance distribution” (SAD) [1–9]. The patterns in the distribution and abundance of species within a biome are central concerns in ecology, as they provide important information about the total species richness, the species-area relation, succession, the likelihood of species extinction under habitat loss, the reserve design, and the processes that allow species to coexist and partition resources [10–13]. This pattern is so fundamental that Sugihara (1980) referred to it as “minimal community structure” [14].

Ecologists have proposed many models describing the SADs in various communities, including the empirical models such as logseries (LSer) [1] and lognormal distribution (LN) [2], and the theoretical models such as geometric series (GS; or niche pre-emption) [15], broken stick (BS) and overlapping niche (ON) [16]. The empirical models have been applied to describe the SAD, while the theoretical ones to describe the curve of the individual number (ranked from most to least, or vice versa) on the species number, i.e., the rank abundance curve (RAC) [11–12,17–20].

Many have conducted a large number of case studies to model the SADs [2,11,14,16, 21–27]. However, little effort has been made to find out the best among these SAD models through the “model selection” method instead of commonly-used zero hypothesis method [28]. But Yin (2005) proposed a “sequential distribution set” by integrating logCauchy (LC) and log-sech (LS) distributions with the widely used lognormal (LN), and modeled the species abundance data from six forest communities (five in lower subtropics and one in tropics) [22,29]. He found that the LC was the best although the three models statistically adequately fitted all the datasets. Recently, Wei et al. (2014) explored the SAD models that are appropriate for six subtropical forests [30]. They used three models, namely the lognormal, logCauchy and log-series, to fit the datasets from six communities, and similarly concluded that the logCauchy model had the best fitness based on the coefficient of determination.

On the other hand, less effort has been made to examine the relationships between the SADs and the succession of communities by using the best model. But Whittaker (1972) evaluated the changes of SAD during the succession from waste farmland to oak forest and concluded that the SAD followed the GS in early stage and the LN in late stage [31]. Yin et al. (2009) selected seven widely representative SAD models to comparatively explore the SAD patterns and dynamics during the natural restoration of degraded hilly grassland in southern China [32]. Their results showed that: (1) the logCauchy distribution best followed the data; (2) the fit of each model to observation was not directly related to successional stage; and (3) the commonness and rarity of species changed with succession. Are these true or false for the forest land, especially in tropics and subtropics?

Here we try to answer the following questions: (1) whether forest communities in different successional stages in lower subtropics follow different SAD models; (2) whether there is a best

SAD model in a widely representative model set; and (3) whether the best model can reveal the intrinsic quantitative features of community structure and dynamics in succession.

Materials and methods

Community survey

We selected two 1-ha (100 m×100 m) permanent forest sites (plots) at Dinghushan Forest Ecosystem Research Station (geographic coordinates: 112°30'39"-112°33'41" E, 23°09'21"-23°11'30" N; see websites: <<http://dhf.cern.ac.cn/content?id=16383>> and <<http://dhf.cern.ac.cn/contentpic?id=36780>>) in Dinghushan Biosphere Reserve, Zhaoqing, Guangdong of China. The two sites established in 1999 are about 2 km apart, and represent two typical forest communities in different successional stages [33]: 70~80-year-old evergreen needle- and broad-leaved mixed forest at Feitianyan, and >400-year-old monsoon evergreen broad-leaved forest at Sanbaofeng. The former is dominated by the tree species *Schima superba*, *Castanopsis chinensis* and *Pinus massoniana*, while the latter by *Cryptocarya concinna*, *C. chinensis* and *Aporosa yunnanensis*.

The abundance (i.e., the number of individuals) of each vascular plant for each plot was investigated by quadrat sampling method. The trees (diameter at breast height, DBH ≥1 cm) were surveyed in the tree layer at 25 adjacent quadrats of 20 m×20 m. Each such quadrat was divided into 16 sub-quadrats (5 m×5 m), one of which was then randomly chosen to record the shrubs and the seedlings of trees (DBH <1 cm, height ≥50 cm) in the shrub layer. Meanwhile, within each of the above-chosen 25 sub-quadrats, a 1 m×1 m quadrat was then randomly selected to count the herbs and the woody seedlings (height <50 cm) in the herb layer.

Model set

Using the method of model selection [28], we assembled the GS, BS, ON and LSer models into the sequential distribution set of the LC+LS+LN to describe the SADs [22,32]. The GS, BS and ON are theoretical models, while the LSer, LC, LS and LN are empirical ones. As mentioned below, the seven models are widely representative with obviously different function types on log scale. They were simultaneously confronted with eight datasets from four layers of two forest communities in different successional stages in South China.

For the model of geometric series (GS, or niche preemption) [31], the abundance proportion of *j*th species is given by

$$p_j = k(1 - k)^{j-1} \tag{1}$$

where $0 < k < 1$; $j = 1, 2, \dots, S$ (the total number of species in a community). This is the RAC model of GS, which is equivalent to the SAD model of truncated hyperbola [13], so the number of species that represents *r* individuals is written as

$$S(r) = \frac{C}{r} \tag{2}$$

where $r \in [r_{\min}, r_{\max}]$; r_{\min} and r_{\max} are the smallest and largest abundance, respectively; and *C* is a constant. On log₂ scale, Eq 2 can be transformed into

$$S(R) = S_m \tag{3}$$

where *R* is the “octave” called by Preston (1948) [2], equivalent to log₂ scale of *r*, i.e., $R = \log_2 r$; *S*(*R*) is the number of species in *R*th octave; *S_m* is a constant.

As to the broken stick (BS) model [16], when the species are arranged from the most common to the rarest, the proportional number of individuals of j th most common species is

$$p_j = \frac{1}{S} \sum_{i=j}^S \frac{1}{i} \tag{4}$$

where $j = 1, 2, \dots, S$. The above RAC-typed BS model and the SAD-typed exponential distribution model are interdeducible [34]. The latter is as follows,

$$S(r) = Ce^{-\alpha r} \tag{5}$$

where $r \in [0, \infty)$; C is a constant. Correspondingly, SAD model on \log_2 scale is given by

$$S(R) = S_m \exp(-2^R \alpha + R \ln 2) \tag{6}$$

where $R \in (-\infty, \infty)$; S_m and α are constants [32].

We can estimate the theoretical total number of species in the whole community, S^* , by extrapolating the curve below the minimal abundance class and measuring the area below the complete (non-truncated) curve [2,21,29]. For Eq 6, we have [32]

$$S^* \approx \frac{S_m}{\alpha \ln 2} \tag{7}$$

In the overlapping niche (ON) model [16], the relative abundance per species follows a linear distribution [35]:

$$f(r) = 2 - 2r \tag{8}$$

where $r \in [0, 1]$. The corresponding SAD model on \log_2 scale was derived [32] as

$$S(R) = S_m \left(1 - \frac{2^R}{N_{(R)}} \right) 2^R \tag{9}$$

where $R \in (-\infty, R_{\max}]$, R_{\max} is the observed largest octave; S_m is a constant; $N_{(R)}$ is the total number of individuals calculated after octave transformation, i.e., $N_{(R)} = \sum 2^R S_{\text{obs}}(R)$, within which $S_{\text{obs}}(R)$ is the observed number of species in R th octave. Similarly, we can get the estimation of S^* [32],

$$S^* \approx \frac{S_m}{2 \ln 2} \tag{10}$$

For the logseries (LSer) model derived by Fisher et al. (1943) [1], the expected number of species having r individuals is given by

$$S_r = C \frac{\alpha^r}{r} \tag{11}$$

where $r = 1, 2, 3, \dots$; $\alpha \in (0, 1)$; C is a constant. Plotkin & Muller-Landau (2002) proposed a corresponding distribution density function [13],

$$f(r) = \frac{\alpha^r}{r \Gamma(0, -\ln \alpha)} \tag{12}$$

where $r \in [1, \infty)$. We can transform the function into the SAD model on \log_2 scale [32],

$$S(R) = S_m \alpha^{2^R} \tag{13}$$

where $R \in [0, \infty)$. The LSer cannot estimate the S^* [21].

The general equation of the lognormal (LN) distribution model is

$$S(R) = S_m \exp[-\alpha^2 (R - R_m)^2] \tag{14}$$

where $R \in (-\infty, \infty)$; R_m is the modal (peak) octave; S_m is the number of species in the R_m th octave (or the height of the curve); α is a constant [2]. Preston (1948) first proposed the estimation of S^* ,

$$S^* \approx \frac{\sqrt{\pi} S_m}{\alpha} \tag{15}$$

Yin (2005) first applied a logCauchy (LC) distribution to community ecology for modeling the observed SADs in forest communities [22]. The LC model on \log_2 scale is

$$S(R) = \frac{S_m}{1 + \alpha^2 (R - R_m)^2} \tag{16}$$

where three parameters, S_m , α , and R_m , are similar to the LN's counterparts. Similarly, the S^* can be estimated by [22,29]

$$S^* \approx \frac{\pi S_m}{\alpha} \tag{17}$$

Yin (2005) also first proposed a log-sech (LS) distribution to describe the species abundance data [22], given by

$$S(R) = S_m \operatorname{sech}[\alpha(R - R_m)] \tag{18}$$

where S_m , α , and R_m are similar to the LN's and LC's counterparts. In the same way, we can deduce the S^* calculating formula [22,29],

$$S^* \approx \frac{\pi S_m}{\alpha} \tag{19}$$

Data analysis

To ensure the comparability of data analyses, we used the same methods for quadrat-sampling and abundance-grouping between two communities, and the same methods for fitting, testing and analyzing among seven models. The abundances for each dataset were grouped by Preston's (1948) octave method [2,36]. The seven models were defined as Eqs (3), (6), (9), (13), (14), (16) and (18) for the GS, BS, ON, LSer, LN, LC, and LS models, respectively. Model parameters were estimated by using Levenberg-Marquardt least squared nonlinear regression of $S(R)$ against R by SPSS 9.0 (SPSS Inc.). We measured the significance of model fitting to data by the chi-square test at 0.05 level. Simultaneously, we evaluated the goodness-of-fit among models by the criterion of the coefficients of determination (i.e., $R_d^2 \geq 0.50$) of the regression [29,37–38], as well as the following Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [39–42],

$$AIC = n \ln(SS_{Residual}) + 2k - n \ln(n) \tag{20}$$

$$BIC = n \ln(SS_{Residual}) + k \ln(n) - n \ln(n) \tag{21}$$

where n is sample size (here the number of observed octaves); k is the number of parameters; and $SS_{Residual}$ is the residual sum of squares. In contrast to the R_d^2 , the lower the AIC or BIC, the better the model.

Results

Species abundance data

The evergreen needle- and broad-leaved mixed forest and the monsoon evergreen broad-leaved forest typically represented two different successional stages in southern subtropical Dinghushan Biosphere Reserve, South China. The latter had more species than the former in each community layer (Table 1). With the succession of the mixed to the broad-leaved, for the species number, the upper layers (pure tree layer and tree layer) increased and the lower layers (shrub layer and herb layer) increased much more; while for the individual number, the upper changed not significantly but the lower more than doubled.

Models' fitting and testing

The observed SADs each showed a left-truncated peak shape (bell shape) with the modal octave $R_m = 1$, except $R_m = 2$ for the tree layer of the broad-leaved forest, in four layers of two forest communities (Fig 1). This suggested that each layer of two communities had many rare species and a few common species. The Chi-square test probability $P(\chi^2)$, the determination coefficient R_d^2 , and two information criteria AIC & BIC generally suggested that the LC model followed all the eight datasets, and moreover, it was the best in the set of seven models (Fig 1; Tables 2–4).

Firstly, we examined the goodness-of-fit of model to data by using chi-square test (Table 2). If the $P(\chi^2) \geq 0.05$, then distribution model confirms with the observation. The LC model followed all the eight datasets, the LN seven, the LS six, the LSer the half, and the GS only two, while the BS and ON models did not all.

The LC also well followed all according to the R_d^2 criterion (Table 2). Although the LSer followed all, its R_d^2 values were lower than the counterparts of LN's, LC's or LS's in most cases. The LN or LS did not follow only one dataset (the herb layer of mixed forest). The order of magnitude in R_d^2 generally was: $LN < LS < LC$. The GS model has no variables and thus its R_d^2 values all were zero. The BS followed only two datasets, while the ON did not follow all due to its negative R_d^2 values in all cases.

Table 1. Descriptive statistics of species abundance data from four layers of two forest communities in Dinghushan Biosphere Reserve, Guangdong of China.

Layer / Community	Total area (m ²)	Total species number, S	Total individual number, N	Min. individual number	Max. individual number
Pure tree, DH	10000	59	2069	1	573
Tree, DH	10000	71	3899	1	773
Shrub, DH	625	39	433	1	83
Herb, DH	25	32	150	1	29
Pure tree, DK	10000	80	2412	1	848
Tree, DK	10000	88	3382	1	995
Shrub, DK	625	94	936	1	238
Herb, DK	25	47	437	1	217

DH: evergreen needle- and broad-leaved mixed forest at Feitianyan; DK: monsoon evergreen broad-leaved forest at Sanbaofeng; Pure tree layer: $H \geq 3$ m; Tree layer: $DBH \geq 1$ cm; Shrub layer: $DBH < 1$ cm and $H \geq 50$ cm; Herb layer: $H < 50$ cm (These are the same elsewhere).

<https://doi.org/10.1371/journal.pone.0196898.t001>

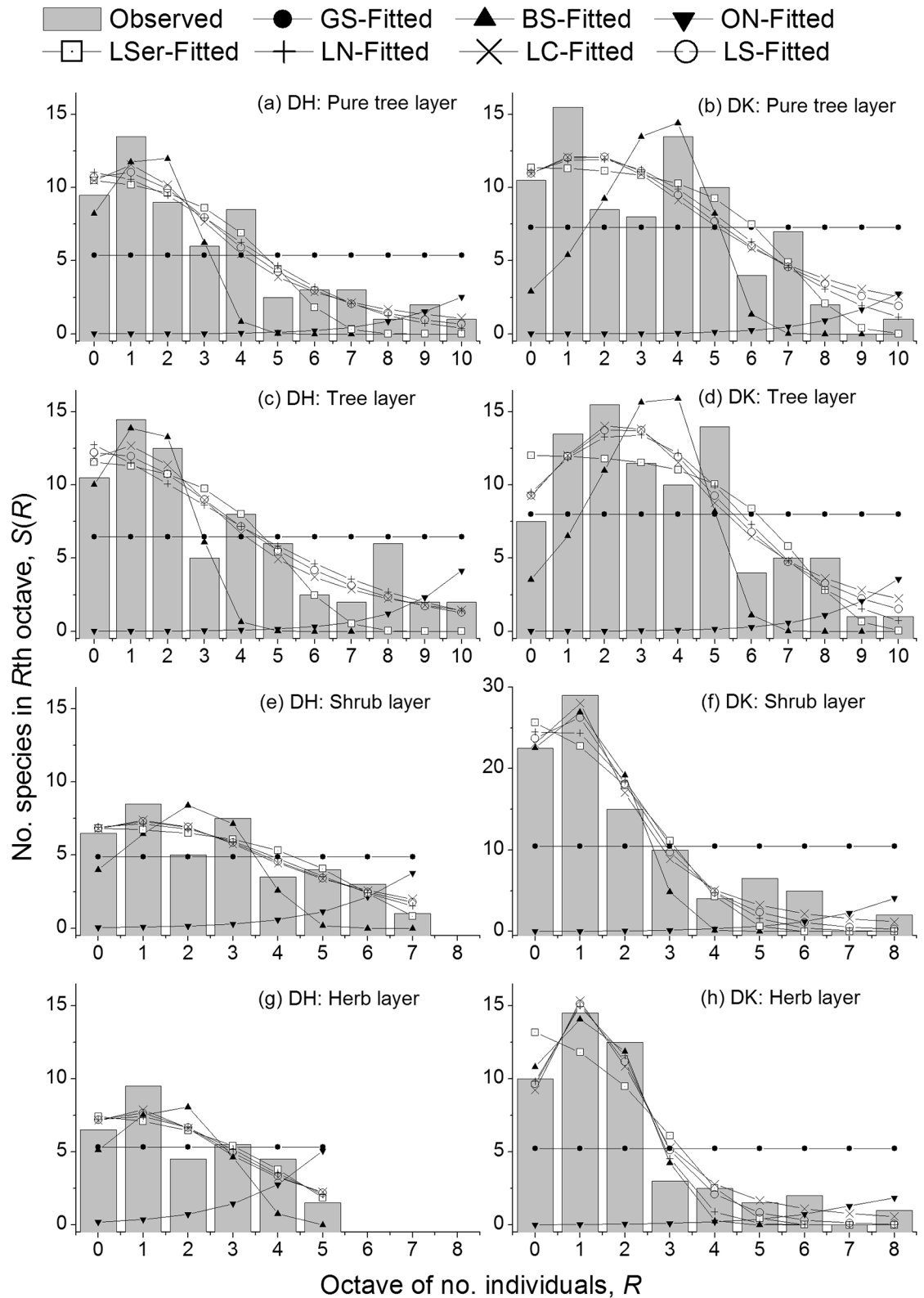


Fig 1. Observed and fitted species-abundance distributions (SADs) in respective four layers within the two communities in Dinghushan Biosphere, Guangdong of China. (Observed: the observed number of species; Fitted: the model-expected number of species. DH: evergreen needle- and broad-leaved mixed forest at Feitianyan; DK: monsoon evergreen broad-leaved forest at Sanbaofeng).

<https://doi.org/10.1371/journal.pone.0196898.g001>

Table 2. Comparisons of the goodness-of-fit of seven SAD models to species abundance data from four layers of two forest communities in Dinghushan Biosphere Reserve, Guangdong of China.

Layer / Community	GS	BS	ON	LSer	LN	LC	LS
Pure tree, DH	0.000	0.351	-1.748	0.746	0.804*	0.845*	0.829*
Tree, DH	0.000	0.175	-2.194	0.532	0.670*	0.736*	0.700*
Shrub, DH	0.000*	-0.270	-3.863	0.700*	0.677*	0.662*	0.672*
Herb, DH	0.000*	-0.270	-3.869	0.598*	0.481*	0.504*	0.493*
Pure tree, DK	0.000	-0.102	-2.218	0.732*	0.668*	0.600*	0.632*
Tree, DK	0.000	0.053	-2.529	0.677*	0.759*	0.745*	0.759*
Shrub, DK	0.000	0.803	-1.232	0.821	0.850	0.949*	0.905
Herb, DK	0.000	0.930	-0.984	0.802	0.929*	0.940*	0.939

Values in this table are coefficients of determination (R_d^2)

*: model statistically significantly agrees with data by using the chi-squared goodness-of-fit test at 0.05 level.

<https://doi.org/10.1371/journal.pone.0196898.t002>

Finally, we compared the seven models by using two information criteria (Tables 3 and 4). For either AIC or BIC, the LSer, LN, LC and LS generally were close to each other, and much lower than GS, BS and ON. In five out of eight cases, the LC was less than the LSer, LN or LS for each of the information criteria, while in other cases, the former was slightly greater than the latter. Similarly to R_d^2 , both AIC and BIC proved that the LC were the best among the seven models.

The fitness of models to the observations was not directly related to the successional stage of community forest. This is because in those scenarios that models were conformed with data, there were the different layers of communities in either early or late successional stages. The GS model by $P(\chi^2)$ followed the datasets only from the shrub and herb layers of mixed forest in early successional stage, while the BS by R_d^2 followed the datasets only from the same two layers of broad-leaved forest in late successional stage. On the other hand, either by $P(\chi^2)$ or R_d^2 , the ON could not describe all the communities either in early or late successional stages, while the LC, LS, LN and LSer could describe most by $P(\chi^2)$ and all by R_d^2 .

Predictions and comparisons of S^*

In the case that the observed SADs were left-truncated distributions, the BS, ON, LN, LC or LS model could be used to calculate the S^* (Table 5). The S^* values predicted by the LN, LC and LS were larger than the corresponding S values, with an exception of the LN-predicted S^* for the herb layer of broad-leaved forest. The BS-predicted S^* values were greater than the corresponding S values only in three out of eight cases, while the ON-predicted S^* values were lower

Table 3. Comparisons of AICs for seven SAD models fitting the species abundance data from four layers of two forest communities in Dinghushan Biosphere Reserve, Guangdong of China.

Layer / Community	GS	BS	ON	LSer	LN	LC	LS
Pure tree, DH	32.5	28.5	43.6	18.2	16.1	13.5	14.6
Tree, DH	33.8	32.5	46.6	26.3	23.1	20.7	22.1
Shrub, DH	15.6	18.3	28.3	6.7	7.9	8.2	8.0
Herb, DH	12.6	14.6	22.1	7.8	9.5	9.3	9.4
Pure tree, DK	36.7	38.6	49.5	23.0	26.1	28.1	27.2
Tree, DK	37.2	37.5	51.1	25.6	23.2	23.8	23.1
Shrub, DK	42.1	28.3	49.4	27.5	26.5	16.8	22.3
Herb, DK	31.7	8.6	37.8	17.9	9.3	7.8	7.9

<https://doi.org/10.1371/journal.pone.0196898.t003>

Table 4. Comparisons of BICs for seven SAD models fitting the species abundance data from four layers of two forest communities in Dinghushan Biosphere Reserve, Guangdong of China.

Layer / Community	GS	BS	ON	LSer	LN	LC	LS
Pure tree, DH	32.9	29.3	44.0	19.0	17.3	14.7	15.8
Tree, DH	34.2	33.3	47.0	27.1	24.3	21.9	23.3
Shrub, DH	15.7	18.5	28.3	6.9	8.1	8.5	8.2
Herb, DH	12.3	14.2	21.8	7.3	8.9	8.7	8.8
Pure tree, DK	37.1	39.4	49.9	23.8	27.3	29.3	28.4
Tree, DK	37.6	38.3	51.5	26.4	24.3	24.9	24.3
Shrub, DK	42.3	28.7	49.6	27.9	27.1	17.3	22.9
Herb, DK	31.9	9.0	38.0	18.3	9.9	8.3	8.5

<https://doi.org/10.1371/journal.pone.0196898.t004>

than the corresponding S values in all the cases. For the estimations of S^* , the LN, LC and LS models were relatively reasonable (i.e., $S^* > S$). However, the LC only had the increasing trend for each of four layers from the mixed to the broad-leaved forest, while both the LN and LS showed much decreasing trend for the tree layer, and slightly decreasing trend for the herb layer. These different performances of five models mentioned above were obviously related to their goodness-of-fits to the data. It was proved from the other side that the LC was the best model.

Comparisons of LC-predicted SAD curves

As mentioned above, the LC model was the best in the set of seven models to agree with all the SAD data from four layers of two communities. So we plotted the probability distribution curves of LC-predicted SADs that were left-truncated (Fig 2). The curves comparatively showed the proportional number of species on the abundance octave for the different layers of different communities.

We first made the comparisons of different layers in the same community (Fig 2). The height of LC-predicted SAD curves of different layers showed different patterns. The order of curve height was herb layer > pure tree layer > tree layer > shrub layer for the mixed forest, while the order was herb layer > shrub layer > tree layer > pure tree layer for the broad-leaved forest. Except that the curve for the herb layer remained the highest, the order of curve height for the other three layers was just reversed from the mixed to the broad-leaved. This suggested that the commonness and rarity distribution of species changed with community succession: generally, the commonness moved up whereas the rarity moved down.

Table 5. Comparisons of the observed and five-SAD-model-expected total numbers of species in four layers of two communities in Dinghushan Biosphere Reserve, Guangdong of China.

Layer / Community	S	S^*_{BS}	S^*_{ON}	S^*_{LN}	S^*_{LC}	S^*_{LS}
Pure tree, DH	59	49.4	7.8	109.3	105.3	92.3
Tree, DH	71	56.9	8.6	236.0	131.5	127.0
Shrub, DH	39	33.4	5.0	62.4	83.7	65.4
Herb, DH	32	32.5	5.6	50.1	65.2	52.4
Pure tree, DK	80	58.2	8.3	116.6	166.8	127.3
Tree, DK	88	65.7	8.7	104.3	146.3	113.7
Shrub, DK	94	105.7	20.8	121.8	136.1	115.9
Herb, DK	47	55.7	10.6	45.8	65.8	51.5

S and S^* : the observed and expected total number of species, respectively.

<https://doi.org/10.1371/journal.pone.0196898.t005>

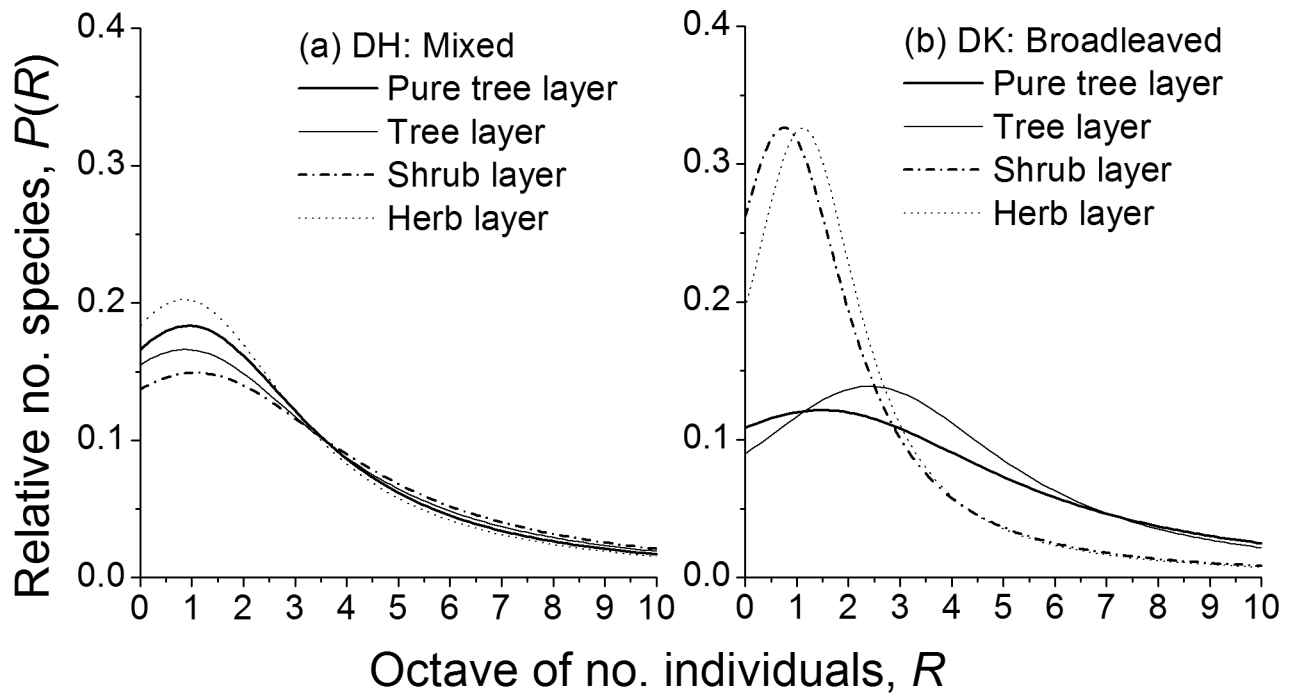


Fig 2. SADs predicted by left-truncated logCauchy (LC) models on octave (\log_2) scale in four layers of two communities in Dinghushan Biosphere Reserve, Guangdong of China. (Area below each distribution curve is set as unity).

<https://doi.org/10.1371/journal.pone.0196898.g002>

The curve height difference of the lower and upper layers between two forest communities also showed an interesting pattern (Fig 2). The curves of four layers were close to one another (i.e., the curves of lower layers were close to the ones of upper layers) in the mixed forest; while the curves of lower layers were much higher than the ones of upper layers in the broad-leaved forest. As a consequence, the height difference of SAD curves between the upper and the lower layers increased with the succession of the mixed to the broad-leaved forest. That is, the difference of species commonness and rarity distribution between the upper and the lower layers increased with succession.

The comparisons of fitted SADs were also made in the same layer between two communities (Fig 2). The height difference of probability curves for the lower layers was distinctly larger than the one for the upper layers between two communities. This suggested that the SADs of lower layers changed more than the ones of the upper layers with the succession. Besides, the order of curve height for the upper and lower layers between two communities was just reverse. The curve of the mixed forest was somewhat higher than the one of the broad-leaved for the upper layers, while the former was much lower than the latter for the lower layers. In other words, in the upper layers, the broad-leaved forest had more proportional common species (the right of curves) and less proportional rare species (the left of curves) than the mixed forest; while in the lower layers, the pattern was opposite. Therefore, with the succession of community, species' commonness moved to the upper layers, but their rarity moved to the lower layers.

Discussion

Selection and mechanism of SAD models

We integrated the GS, BS, ON and LSer into the “sequential distribution set” composed of the LC, LS and LN [29], and produced a larger set of seven models by using model selection method

[28]. The new set covers the empirical (statistical) and theoretical (niche) models and has very different function forms, so it has stronger modeling capability. Nevertheless, such a set of SAD models can be extended. As to the fitness order of models, the study was very similar to the previous studies [29,32]. This further proved that the LC model was the best in the set of either three or seven models. Yin et al. (2013) also fitted a set of 12 statistical distribution models to the datasets from four body size indicators (canopy diameter, breast-height diameter, ground diameter and tree height) and their six derived area and volume indicators of *Castanopsis hystrix* in Guangzhou, South China, and stated that the LC distribution was one of the best models which were commonly accepted [43]. So we conclude that the logCauchy model has the widest applicability in population and community ecology.

Ecologists proposed some mechanisms to interpret LN-typed SADs. The central limit theorem of statistics predicts that the LN distribution should produce when many variables interact multiplicatively [16,44–45]. Sugihara (1980) stated that the LN results from the species' subdividing niche space within a community [3,13,46]. Engen & Lande (1996) derived the SAD of LN type from population dynamics models [47]. Hubbell's (2001) neutral theory suggested that the SAD of LN-like shape is a particular case of his zero-sum multinomial distribution [48]. No matter what its origin is, the LN is important because it can predict the distribution of individual abundance among species (i.e., SAD, and thus S^*), and has provided other valuable insights into the community organization such as species-area relationship and community succession [10–12,49–50]. The LC distribution should be also suitable for the scenarios of the LN because it has some characteristics similar to the latter [22,32].

Pattern and dynamics of SAD

The SADs followed the GS model only for the shrub layer and herb layers (i.e., the lower layers) of the mixed forest in early successional stage, while they followed the LN (as well as the LC) for each layer of two forests in different successional stages. That is, whether for seven models to follow the observed was not directly related to the successional stage of community. This was somewhat different from Whittaker's (1965) study that indicated that the SAD changed from the GS to the LN with the succession in an old field [15], but was similar to the results on the pattern and dynamics of SAD during the natural restoration of degraded hilly grassland in southern China [32].

For above reasons, we cannot qualitatively identify the community successional stage according to the SAD model type, and vice versa. However, we can quantitatively examine the relationship between the SAD pattern and community succession by using a universal model (here the LC). The LC-predicted SAD curves indicated that with the succession of forest community, the commonness of species moved up while the rarity of species moved down, and the difference of the commonness-rarity distribution of species increased between the upper and the lower layers. These results are confirmed with the study by Yin et al. (2009) [32]. With the development of community, the number of tree species increase and their spreading layers become higher and more, so the common species in the upper layers increase. Meanwhile, the understory plants (especially the herbs and the seedlings of trees and shrubs intolerant of shade) decrease with increasing canopy density, so the species rarity increases in the lower layers of community.

Implications for studying SAD

He and Legendre (2002) stated that the area, abundance, richness, and spatial distribution of species have been central components of community ecology [51]. The quantitative studies on the structure and dynamics of a natural forest community can help solve some practical

problems such as species selection and configuration in the reconstruction of forest vegetation [22,52]. This is because that the SAD can reveal the commonness-rarity of species, the proportional individual number of every species, and the distribution dynamics of different layers in natural communities. Our modeling results showed that the logCauchy model had the widest applicability and the best fitness to all the datasets among the seven models, and could best interpret the SAD patterns and dynamics of communities and their different layers in the succession of forests. These results will quantitatively provide feasible guidance for the reconstruction or improvement of public forests and the restoration of degraded vegetation. This study is also expected to enrich the theories in community ecology, macroecology, biogeography and restoration ecology.

Conclusion

The SADs all showed left-truncated peak shape on log scale for the four layers of two forest communities in Dinghushan Biosphere Reserve, Guangdong of China. The model selection indicated that the LC model was the best among the set of seven models. We also found whether for the SAD models to follow the data was not directly related to the successional stage of forest community. Although we cannot qualitatively identify the successional stage of a community according to the type of SAD models and vice versa, we can quantitatively examine the changes of the SAD of community with succession through a commonly accepted model, the LC. The LC-predicted SAD curve showed that the commonness of species moved up while the rarity of species moved down with the succession of forest community. This is because that the broad-leaved forest had more proportional common species and less proportional rare species than the mixed forest in the upper layers, while the case was reverse in the lower layers. Moreover, the difference of the SADs between the upper and the lower layers increased with succession.

Supporting information

S1 File. The dataset necessary to replicate our study findings.
(XLS)

Acknowledgments

We thank Hai Ren, Qian-Mei Zhang, Guo-Yi Zhou, Zhong-Liang Huang, Shi-Zhong Liu, Wan-Hui Ye, Qing-Fa Yu, You-Chang Zhang, Xiao-Yong Zhou, Yuan-Qing Jiang, Ding-Sheng Mo, and many others in South China Botanical Garden (SCBG), the Chinese Academy of Sciences for their help in field work. We thank Dr. Prof. Qinfeng Guo within Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, USA and Dr. Prof. Shao-Lin Peng within Sun Yat-Sen University, China for their constructive suggestions. We also thank the Dinghushan Forest Ecosystem Research Station of SCBG in Dinghushan Biosphere Reserve, Guangdong, China for supporting the field investigations.

Author Contributions

Conceptualization: Zuo-Yun Yin, Shao-Ming Luo.

Data curation: Zuo-Yun Yin, Lu Zeng, Wei Guo.

Formal analysis: Zuo-Yun Yin, Shao-Ming Luo, Xiao He, Wei Guo, Bailian Li.

Funding acquisition: Zuo-Yun Yin, Lu Zeng, Ping Chen.

Investigation: Zuo-Yun Yin, Wei Guo.

Methodology: Zuo-Yun Yin, Shao-Ming Luo, Xiao He, Wei Guo, Bailian Li.

Project administration: Zuo-Yun Yin, Lu Zeng, Shao-Ming Luo, Ping Chen, Wei Guo.

Resources: Zuo-Yun Yin, Lu Zeng, Shao-Ming Luo, Ping Chen, Wei Guo.

Software: Zuo-Yun Yin, Shao-Ming Luo, Xiao He.

Supervision: Lu Zeng, Shao-Ming Luo, Ping Chen, Wei Guo.

Validation: Zuo-Yun Yin, Xiao He.

Visualization: Zuo-Yun Yin.

Writing – original draft: Zuo-Yun Yin.

Writing – review & editing: Zuo-Yun Yin, Lu Zeng, Shao-Ming Luo, Ping Chen, Xiao He, Wei Guo, Bailian Li.

References

1. Fisher RA, Corbet AS, Williams CB. The relation between the number of species and the number of individuals in a random sample from an animal population. *Journal of Animal Ecology*. 1943; 12: 42–58.
2. Preston FW. The commonness and rarity of species. *Ecology*. 1948; 29: 254–283.
3. Molles MC. *Ecology: Concepts and Applications*. New York: McGraw-Hill Science Engineering; 1999.
4. Peng SL, Yin ZY, Ren H, Guo QF. Advances in research on the species-abundance relationship models in multi-species collection. *Acta Ecologica Sinica*. 2003; 23(8): 1590–1605.
5. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*. 2007; 10(10): 995–1015. <https://doi.org/10.1111/j.1461-0248.2007.01094.x> PMID: 17845298
6. Jost L, Chao A, and Chazdon RL. Compositional similarity and β (beta) diversity. In: Magurran AE, McGill BJ, editors. *Biological Diversity Frontiers in Measurement and Assessment*. New York: Oxford University Press Inc. 2011; PP 66–335.
7. Rocchini D and Neteler M. Spectral rank–abundance for measuring landscape diversity. *International Journal of Remote Sensing*. 2012; 33(14): 4458–4470. <https://doi.org/10.1080/01431161.2011.648286>
8. Guo QF. Central-marginal population dynamics in species invasion. *Frontiers in Ecology and Evolution*. 2014; 2: 1–17. <https://doi.org/10.3389/fevo.2014.00023>
9. Ferreri L, Giacobini M, Bajardi P, Bertolotti L, Bolzoni L, Tagliapietra V, et al. Pattern of Tick Aggregation on Mice: Larger Than Expected Distribution Tail Enhances the Spread of Tick-Borne Pathogens. *PLoS Computational Biology*. 2014; 10(11): 1–12, e1003931. <https://doi.org/10.1371/journal.pcbi.1003931> PMID: 25393293
10. Bazzaz FA. Plant species diversity in old-field successional ecosystems in southern Illinois. *Ecology*. 1975; 56: 485–488.
11. Aoki I. Diversity and rank-abundance relationship concerning biotic compartments. *Ecological Modelling*. 1995; 82(1): 21–26.
12. Jackson CR, Churchill PF, Roden EE. Successional changes in bacterial assemblage structure during epilithic biofilm development. *Ecology*. 2001; 82(2): 555–566.
13. Plotkin JB, Muller-Landau HC. Sampling the species composition of a landscape. *Ecology*. 2002; 83(12): 3344–3356.
14. Sugihara G. Minimal community structure: an explanation of species abundance pattern. *American Naturalist*. 1980; 116: 770–787. <https://doi.org/10.1086/283669> PMID: 29513556
15. Whittaker RH. Dominance and diversity in land plant communities. *Science*. 1965; 147: 250–260. <https://doi.org/10.1126/science.147.3655.250> PMID: 17788203
16. MacArthur RH. On the relative abundance of species. *American Naturalist*. 1960; 94: 25–36.
17. Li BL. Fractal dimensions. In: El-Shaarawi AH, Piegorsch WW, editors. *Encyclopedia of Environmetrics*. Chichester: John Wiley & Sons; 2002. pp. 821–825.
18. Zhang JT. *Quantitative Ecology*. Beijing: Science Press; 2004. <https://doi.org/10.1186/1472-6785-4-8>

19. Eliazar II, Cohen MH. Topography of chance. *Physical Review E*. 2013; 88(5): 052104.
20. Eliazar II. Power-law and exponential rank distributions: A panoramic Gibbsian perspective. *Annals of Physics*. 2015; 355: 322–337.
21. Kempton RA, Taylor LR. Log-series and log-normal parameters as diversity discriminants for the Lepidoptera. *Journal of Animal Ecology*. 1974; 43: 381–399.
22. Yin ZY. Modeling on species abundance and distribution patterns in plant communities. Doctoral dissertation, South China Botanical Garden and Graduate School, the Chinese Academy of Sciences. 2005.
23. Loehle C. Species abundance distributions result from body size-energetic relationships. *Ecology*. 2006; 87: 2221–2226. [https://doi.org/10.1890/0012-9658\(2006\)87%5B2221:SADRFB%5D2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87%5B2221:SADRFB%5D2.0.CO;2) PMID: 16995622
24. Forster MA and Warton DI. A metacommunity-scale comparison of species-abundance distribution models for plant communities of eastern Australia. *Ecography*. 2007; 30: 449–458. <https://doi.org/10.1111/j.2007.0906-7590.04899.x>
25. Šizling AL, Storch D, Reif J, Gaston KJ. Invariance in species-abundance distributions. *Theoretical Ecology*. 2009; 2: 89–103. <https://doi.org/10.1007/s12080-008-0031-3>
26. Zhang F, Hong B, Li YM, Chen ZJ, Zhang SL, Zhao HY. Species-area and species-abundance relationships of arthropod community in various vegetation restoration areas in Zhifanggou watershed, Shaanxi Province of Northwest China. *The Journal of Applied Ecology*. 2013; 24(2): 511–516.
27. Scott D. Analogue Particle and Random Packing Models of Ecological Community Formation. *Journal of Plant Studies*. 2014; 3(2): 58–90. <https://doi.org/10.5539/jps.v3n2p58>
28. Johnson JB, Omland KS. Model selection in ecology and evolution. *Trends in Ecology and Evolution*. 2004; 19(2): 101–108. <https://doi.org/10.1016/j.tree.2003.10.013> PMID: 16701236
29. Yin ZY, Peng SL, Ren H, Guo QF, Chen ZH. LogCauchy, log-sech and lognormal distributions of species abundances in forest communities. *Ecological Modelling*. 2005a; 184(2–4): 329–340.
30. Wei SG, Li L, Chen ZC, Lian JY, Lin GJ, Huang ZL, Yin ZY. Which Models Are Appropriate for Six Sub-tropical Forests: Species-Area and Species-Abundance Models. *PLOS ONE*. 2014; 9(4): 1–9.
31. Whittaker RH. Evolution and measurement of species diversity. *Taxon*. 1972; 21: 213–351.
32. Yin ZY, Ren H, Peng SL, Guo QF, Zeng LH, He X. Dynamics and modeling of species abundance distribution during natural restoration of degraded hilly grassland in south China. *Ecology and Environmental Sciences*. 2009; 18(1): 222–228.
33. Zhang QM, Huang ZL, Liu SZ, Zhou GY. The community structure of coniferous and broad-leaved mixed forest in the Dinghushan. In: Zhou GY, Zhang QM, Wen DZ, Peng CL, Zhang DQ (Dinghushan Forest Ecosystem Research Station, DFERS), editors. *Tropical and Subtropical Forest Ecosystem* (Vol. 9); 2002. Beijing: China Environmental Science Press. pp. 18–27.
34. Longuet-Higgins MS. On the Shannon-Weaver index of diversity, in relation to the distribution of species in bird censuses. *Theoretical Population Biology*. 1971; 2: 271–289. PMID: 5162689
35. Pielou EC, Arnason AN. Correction to one of MacArthur's species-abundance formulas. *Science*. 1966; 151: 592. PMID: 5903588
36. McGill BJ. A test of the unified neutral theory of biodiversity. *Nature*. 2003a; 422: 881–885.
37. Yin ZY, Liao WB. Studies on lognormal distribution patterns of species abundance of south subtropical forest community, China. *Guihaia*. 1999; 19(3): 221–224.
38. Yin ZY, Ren H, Zhang QM, Peng SL, Guo QF, Zhou GY. Species abundance in a forest community in South China: a case of Poisson lognormal distribution. *Journal of Integrative Plant Biology*. 2005b; 47(7): 801–810.
39. Akaike H. Information theory and an extension of maximum likelihood principle. In: Petrov BN, Csaki F, editors. *Proceedings of the Second International Symposium of Information Theory*. Budapest, Hungary: Akademiai Kiado. 1973; PP. 267–281.
40. Schwartz G. Estimating the dimension of a model. *Annals of Statistics*. 1978; 6: 461–464.
41. Quinn GP, Keough MJ. *Experimental Design and Data Analysis for Biologists*. Cambridge: Cambridge University Press. 2002; P144.
42. Jiang RY. *Introduction to Quality and Reliability Engineering*: Beijing: Science Press. 2015; PP 79–87.
43. Yin ZY, Zeng LH, He BX, Lian HM, Zhang Q, Cai YL, Chen YQ, Lan YQ. Statistical distribution models of body sizes of young *Castanopsis hystrix* in an urban landscape forest. *Ecology and Environmental Sciences*. 2013; 22(2): 189–198.
44. May RM. Patterns of species abundance and diversity. In: Cody ML, Diamond JM, editors. *Ecology and Evolution of Communities*. Cambridge, MA: The Belknap Press of Harvard University Press; 1975. pp. 81–120.

45. McGill BJ. Strong and weak tests of macroecological theory. *Oikos*. 2003b; 102(3): 679–685.
46. Krebs CJ. *Ecology: the Experimental Analysis of Distribution and Abundance*. 5th ed. San Francisco, CA: Benjamin Cummings; 2001.
47. Engen S, Lande R. Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences*. 1996; 132: 169–183. PMID: [8714412](#)
48. Hubbell SP. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press; 2001.
49. Preston FW. The canonical distribution of commonness and rarity. *Ecology*. 1962; 43: 185–215, 410–432.
50. Miller RI, Wiegert RG. Documenting completeness, species-area relations, and the species-abundance distribution of a regional flora. *Ecology*. 1989; 70(1): 16–22.
51. He F, Legendre P. Species diversity patterns derived from species-area models. *Ecology*. 2002; 83(5): 1185–1198.
52. Zeng LH, Yin ZY. *The Management of Secondary Tropical Forests*. Guangzhou: Guangdong Science and Technology Press, Guangdong Publishing Group; 2013.