

# *Theoretical Approach for How Species Abundance Distributions Change Across Scales\**

*(Moments as a Function of Area)*

Saeid Alirezazadeh, Luís Borda-de-Água  
CIBIO, InBIO, CEABN  
Portugal

saeid.alirezazadeh@gmail.com, lbagua@gmail.com

Francisco Dionísio  
FCUL, CE3C  
Portugal  
francisco.dionisio@gmail.com

Pedro Cardoso  
FMNH, UofH  
Finland  
pedro.cardoso@helsinki.fi

Paulo Borges, Rosalina Gabriel  
CE3C/GBA, UAc  
Portugal

paulo.av.borges@uac.pt, rosalina.ma.gabriel@uac.pt

Henrique M. Pereira  
iDiv, CIBIO, InBIO, CEABN  
Germany  
hpereira@idiv.de

**Abstract**—Which distribution commonly used to fit a species abundance distribution gives the best fit across several area sizes? We approach this problem by looking at the moments of the distribution and compare these with those the distribution based on empirical data on tree species collected in Barro Colorado Island, Panama. We conclude that no single distribution gives the best fit at all scales.

**Keywords**—moments; species abundance distribution; scale; Barro Colorado Island

## I. INTRODUCTION

Any attempt to find an ecological pattern across scales is important not only for ecologists but also for managers. It allows us to describe the pattern at observable scales and to make predictions for larger scales. The species-area relationship is one of the oldest pattern in ecology involving a scaling relationship, and it is commonly used by ecologists. It describes how the number of species (species richness) changes as a function of area and it is formulated as a power law.

Finding a pattern for species abundance distributions (hereafter SAD) at a fixed scale has been a central topic in ecology. SADs describe how many species in a community have a given number of individuals. Fisher et al. [1] suggested that the logseries is the theoretical distribution for the relative abundance of species. Preston [2], by looking at SADs at different scales, proposed, instead, the lognormal as the limiting distribution. Bulmer [3] showed that the Poisson lognormal is a better alternative to the lognormal. More recently, Engen and

Lande [4] also considered the compounded Poisson Gamma distribution. May [5] stated that the logseries may be viewed as the distribution characteristic of relatively simple communities whose dynamics is dominated by some single factor, but, on the other hand, if the environment is randomly fluctuating, or if several factors become significant the central limit theorem will produce the lognormal distribution.

Although most of the work on SADs has been done by looking at a fixed scale, there are exceptions. Hubbell [6] developed the neutral theory that seeks to predict SADs in space and time across scales. This theory was called neutral because it assumed all individuals in a community to be equivalent in their rates of reproduction, mortality, dispersal and speciation. Two of the speciation modes considered by Hubbell, the point mutation and the fission modes, led to SADs that changed differently depending on the scale. Although the fission mode led to the same type of distribution at different scales, the point mutation mode predicted the zero-sum multinomial distribution (a distribution introduced by Hubbell [6]) at small scales (the local community) and predicted the logseries distribution at large scales (the metacommunity). Therefore, neutral theory acknowledges that the SAD changes across scales.

Typically, empirical SADs come from samples that represent only a small part of the community, those sizes that are practical (or economical) to obtain. If we could find a pattern for how SADs scale with area (or other parameter), this could enable us to predict the distributions for larger areas, at least within some reasonable scales, likely to be dictated by the

---

This work was supported by Fundação para a Ciência e Tecnologia (PTDC/BIA-BIC/5558/2014).

characteristics, such as the homogeneity of the habitats, of the landscape where the community exists.

*Moment* functions are used in image analysis and related applications, such as pattern recognition, object identification, template matching, and pose estimation (see [7, 8, 9]). However, their application to the study of SADs have only recently been attempted [10, 11]. Here, we define moments,  $M_n$  (often called raw moments), of a sample  $x_1, x_2, \dots, x_S$  as

$$M_n = \frac{1}{S} \sum_{i=1}^S x_i^n, \quad (1)$$

where  $n$  defines the order of the moment. For instance, when  $n = 1$  we obtain the mean. Moments are important because knowing them is enough to reconstruct the probability density function [12].

In this work we are interested in the scaling properties of the SAD as a function of area, though others variables, such as, diameter at breast height (dbh) could have been used. Therefore, given an area  $A$ , we determine the SAD for several subareas  $A' \leq A$  and see how the distributions change as a function of area size. In addition, we asked the following question: among some commonly used distribution, which gives the best fit to empirical data at different scales.

We used the following distributions: logseries, lognormal, Poisson lognormal, and Poisson gamma [13, 14]. For some distributions we have simple analytical formulas for the moments, in other cases the analytical expressions have to be solved using numerical methods. Finally, we compare the results obtained with the distributions with those obtained empirically from real data.

In order to estimate the moments from the distributions as a function of area, we needed to make some assumption for how the number of individuals and the number,  $N$ , of species,  $S$ , scale with area. For the number of individuals we assumed that it changes linearly with area,  $A$ , such as,  $N = \rho A$  [6]. For the number of species we assumed two different relationships: (i) the number of species has a power law relationship with the area,  $A$ , or, equivalently,

$$\log(S) = \log(c) + z \log(A), \quad (2)$$

where  $c$  and  $z$  are constants, and (ii) it has a logarithmic relationship

$$S = c + z \log(A), \quad (3)$$

where  $c$  and  $z$  are constants. We assumed these two different expressions because they both give a very good fitting to a large number of data sets and are often used [15].

We exemplify our methods using data on tropical tree species from a 50ha plot in Barro Colorado Island (BCI), Panama, using all individuals larger than 10 cm dbh.

## II. PRELIMINARIES

Because in a typical sample the number of individuals spans several orders of magnitude, it is more convenient to work with their logarithmic values [5]. Let  $x_i$  be the log of base 2 of the total number of individuals of the species  $i$  and let  $S$  be the total

number of species in the sample. Then, following eq. 1 the (*raw*) *moment of order  $n$*  is

$$M_n = \frac{1}{S} \sum_{i=1}^S x_i^n,$$

or, if the  $x_i$  occurs with the proportion  $f(x_i)$  in the sample,

$$M_n = \sum_{i=1}^k x_i^n f(x_i), \quad (4)$$

where  $k$  is the number of classes.

In Fig.1, using BCI data, we plot the moments up to order ten as a function of area size in double logarithmic scales, starting from area size  $65 \times 130 \text{ m}^2$ . Notice that the curves are approximately linear, implying that the moments,  $M_n$ , as a function of area,  $A$ , are well approximated by power laws.

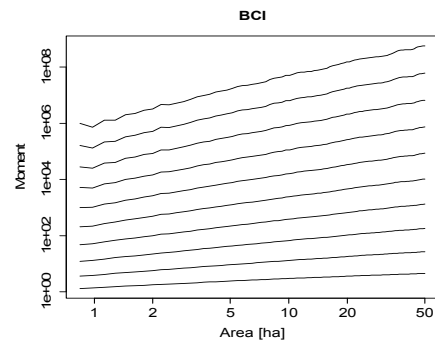


Fig. 1. Moments, estimated from the BCI data, of order 1 up to 10 as a function of area from 0.845 ha0 up to 50 ha. Notice that the double logarithmic scales.

## III. MOMENTS AS A FUNCTION OF SCALE

Here, for each area size we assume that the SAD follows one of the following distributions: logseries, lognormal, Poisson lognormal, and Poisson gamma. For each sub-area we estimate the parameters of the distribution and calculate the moments for different orders. Here we are mostly interested in the log-transformed value of the moments as a function of log of the area size. In the following, the formulas for moments of the distributions were obtained from [16] for the logseries, from [17] for the lognormal, from [18] for the *mixed Poisson distributions* together with [19].

### A. Logseries

Using the expression of the logseries and the above eq. (4), the moments of order  $n$  of a sub-area of size  $A'$  of the total area  $A$  can be obtained from

$$M_n(A') = \sum_{x=0}^{N'} x^n \frac{1}{-\ln(1 - \theta(A'))} \theta(A')^x, \quad (5)$$

where  $\theta(A')$  is the parameter of logseries,  $N' = \lfloor \log_2 N \rfloor + 1$ ,  $N$  is the total number of individuals in the sub-area  $A'$ , and  $\lfloor x \rfloor$  stands for truncation of  $x$ .

In Fig.2, we show in double logarithmic plots the moments, as a function of area using formula (5), up to order 10, in plot (a) assuming  $S = cA^z$  and in plot (b)  $S(A) = c + z \log(A)$ . Observe

that in both cases, for larger area sizes the curves are almost straight, implying a power law relationship.

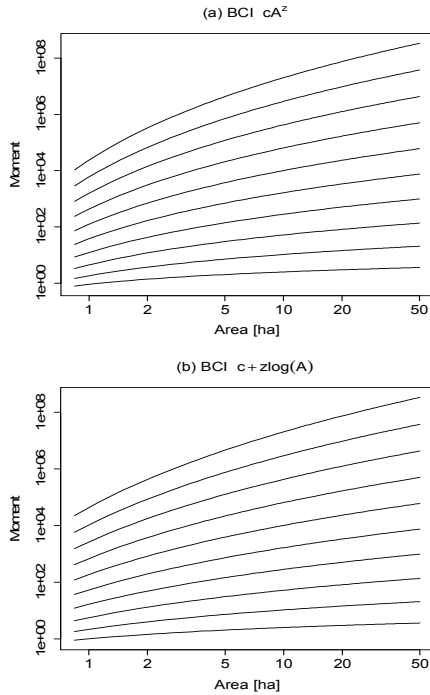


Fig. 2. Moments of order 1 up to 10 as a function of area from 0.845 ha up to 50 ha in double logarithmic scales for the logseries distribution assuming in (a)  $S(A) = cA^z$  and in (b)  $S(A) = c + z\log(A)$ .

### B. Lognormal

The moments of order  $n$  of a sub-area of size  $A'$  of the total area  $A$  can be obtained from the following formula

$$M_n(\mu(A'), \sigma(A')) = \frac{\mu(A')^n + \sum_{i=1}^n \binom{n}{i} \mu(A')^{n-i} \begin{cases} 0 & , i \text{ is odd} \\ \sigma(A')^i (i-1)!! & , i \text{ is even} \end{cases}}{\ln(2)^n}$$

where  $\mu(A')$  and  $\sigma(A')$  are the parameters of the lognormal in the area  $A'$  and the double factorial is defined as follows:

$$n!! = \begin{cases} \prod_{k=1}^{\frac{n}{2}} (2k) & , n \text{ is odd} \\ \prod_{k=1}^{\frac{n+1}{2}} (2k-1) & , n \text{ is even.} \end{cases}$$

In Fig.3, we plot the moments calculated using the preceding formula, up to order ten, as a function of area size in double logarithmic scales. Again, in Fig.3, we assumed two expressions for the species area relationship, eq. (2) and (3). As we can see, a linear pattern only occurs for larger areas, and for small areas the curvature is more pronounced than it is for the logseries. Note that in the preceding formulas we may obtain negative values for the parameter  $\sigma^2(A')$  implying that for those sub-areas the lognormal cannot be used.

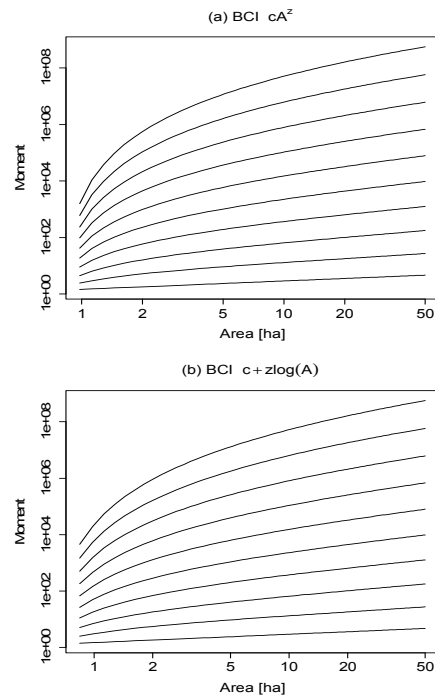


Fig. 3. Moments of order 1 up to 10 as a function of area from 0.845 ha up to 50 ha in double logarithmic scales for the lognormal distribution assuming in (a)  $S(A) = cA^z$  and in (b)  $S(A) = c + z\log(A)$ .

### C. Poisson Lognormal

The moments of order  $n$  of a sub-area of size  $A'$  of the total area  $A$  can be obtained from the following formula

$$M_n(\mu(A'), \sigma(A')) = \frac{\sum_{j=1}^n S(n,j) M'_j(\mu(A'), \sigma(A'))}{\ln(2)^n}$$

where  $S(n,k)$  denotes the Stirling numbers of the second kind, defined as

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n$$

$\mu(A')$  and  $\sigma(A')$  are the parameters of the Poisson lognormal distribution, and

$$M'_j(\mu(A'), \sigma(A')) = \mu(A')^j + \sum_{i=1}^j \binom{j}{i} \mu(A')^{j-i} \begin{cases} 0 & , i \text{ is odd} \\ \sigma(A')^i (i-1)!! & , i \text{ is even} \end{cases}$$

In Fig.4 we plot the moments, calculated by the preceding formula, up to order ten, as a function of the area size in double logarithmic scales.

Note that in the case of Poisson lognormal we do not consider different species area relationship, because parameters can be extracted directly from the first and the second moments and there is no need for extra assumptions.

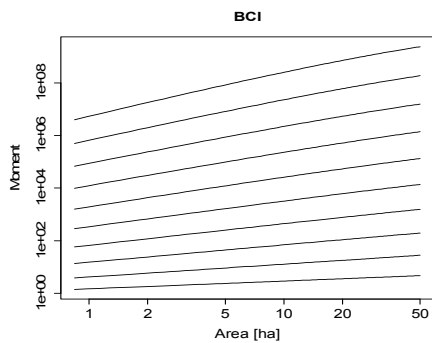


Fig. 4. Moments of order 1 up to 10 as a function of area from 0.845 ha up to 50 ha in double logarithmic scales for the Poisson lognormal distribution.

#### D. Poisson Gamma

The moments of order  $n$  of a sub-area of size  $A'$  of the total area  $A$  can be obtained from the following formula

$$M_n(A') = \sum_{i=1}^n B_{i,n}(c_1(A'), \dots, c_{n-i+1}(A')),$$

where  $c_i$ 's are the  $i$ -th cumulant around the mean and  $B_{i,n}(x_1, \dots, x_{n-i+1})$  is the incomplete (or partial) Bell polynomial which is defined as follows:

$$B_{i,n}(x_1, \dots, x_{n-i+1}) = \sum \frac{n!}{j_1! j_2! \dots j_{n-i+1}!} \left(\frac{x_1}{1!}\right)^{j_1} \left(\frac{x_2}{2!}\right)^{j_2} \dots \left(\frac{x_{n-i+1}}{(n-i+1)!}\right)^{j_{n-i+1}},$$

where the sum is taken over all sequences  $j_1, \dots, j_{n-i+1}$  of non-negative integers such that these two conditions are satisfied:

$$j_1 + j_2 + \dots + j_{n-i+1} = i,$$

$$j_1 + 2j_2 + 3j_3 + \dots + (n-i+1)j_{n-i+1} = n,$$

and  $c_i(A')$ 's can be obtained from the following formula:

$$c_i(A') = \frac{\partial^i D(\eta_1(A'), \eta_2(A'))}{\partial \eta_1(A')^i},$$

where

$$\eta_1(A') = \alpha(A') - 1, \quad \eta_2(A') = -\beta(A'),$$

and

$D(\eta_1(A'), \eta_2(A')) = \ln(\Gamma(\eta_1(A') + 1)) - (\eta_1(A') + 1)\ln(-\eta_2(A'))$ , where  $\alpha(A')$  and  $\beta(A')$  are the parameters of Gamma distribution of a sub-area  $A'$ .

In Fig.5 we plot the moments, calculated by the preceding formula, up to order ten, as a function of the area size in double logarithmic scales, and, as before, we assumed two expressions for the species area relationship, eq. (2) and (3). Similarly to the other distributions, the moments exhibit an almost linear relationship.

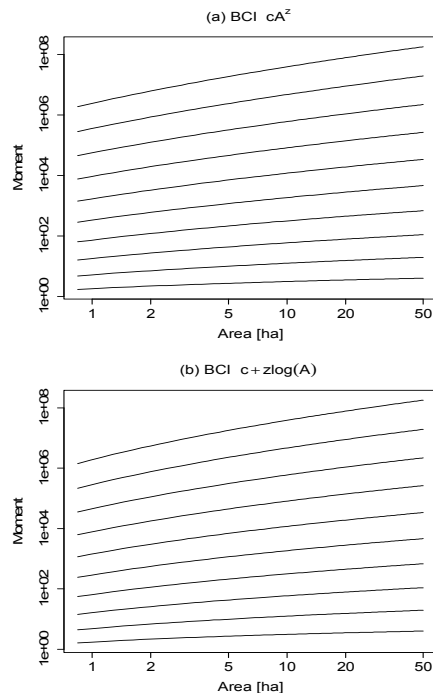


Fig. 5. Moments of order 1 up to 10 as a function of area from 0.845 ha up to 50 ha in double logarithmic scales for the Poisson gamma distribution assuming in (a)  $S(A) = cA^z$  and in (b)  $S(A) = c + z\log(A)$ .

#### IV. COMPARING RESULTS

We now compare the change of the moments as a function of area using the preceding distributions with the BCI data. To better visual the results we show only the moment of order 10, Fig.6. A brief inspection of this figure shows that the Poisson gamma (PG) is close to the empirical data for small area sizes, but that for larger areas the lognormal (LN) provides a better fit. Surprisingly, the logseries provides a poor fitting even for small area sizes.

In order to present these results in a different way, we developed an approach based on the following result. If the moments obtained from a distribution provide a good fit to the moments empirically observed, then if we plot the empirical moments as a function of the distributions' moments, their points should lie in a straight line going through the origin and with slope 1.

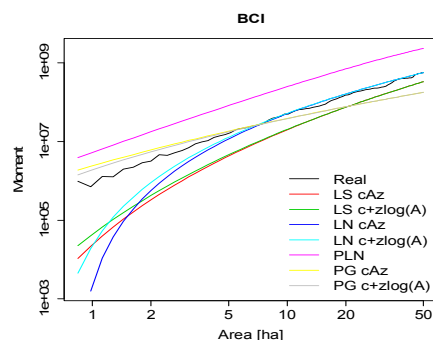


Fig. 6. Comparing the moment of order 10 of the real data (black line) with the moments estimated using the distributions.

Fig. 7 shows the slopes of linear regression of  $m$  by  $n$ . Since the best fitting appears when the slope is closed to 1 we only plot the slopes in the interval 0 and 3.5. In accordance with our previous observations, the Poisson lognormal is not a good fit to our data, but Poisson Gamma distribution is a good fitting for small area sizes. When area increases the lognormal provides the best fitting.

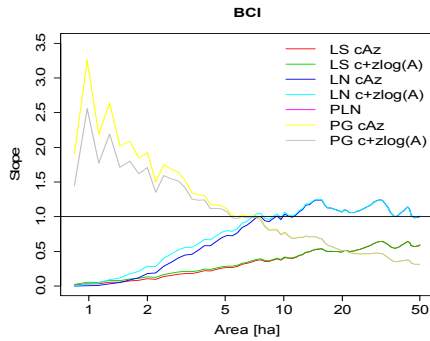


Fig. 7. Comparison of the slopes of moments calculated from the distributions as a function of moments obtained empirically for areas sizes from 0.845 ha up to 50 ha. The closer the slope is to one the better the fit provided by the distribution.

### A. Comparing by different sub-areas

In the previous section, we argued that the closer the slope of linear regression of the empirical moments as a function of those of the distributions were to one, the better the fit. Here, we show the same results explicitly for four different areas 0.98, 12.5, 25.2 and 50 ha, see Fig. 8. According to our previous discussion the distribution that provides the best fit is the one closer to the line going through zero and has slope one (the black line in Fig. 8).

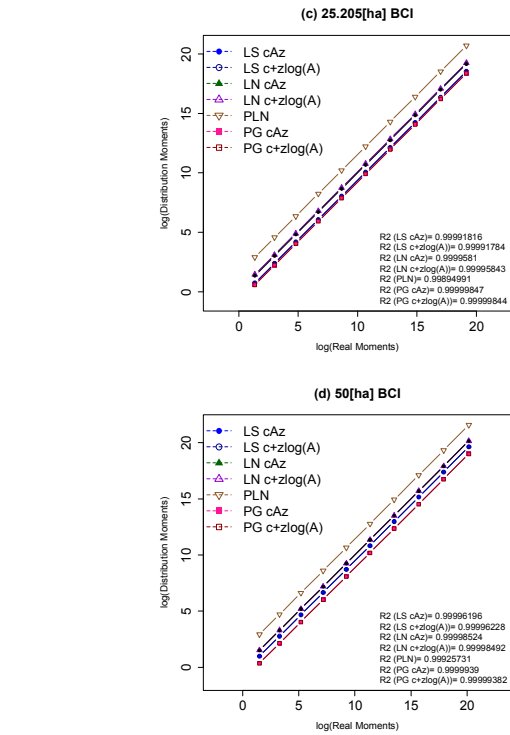
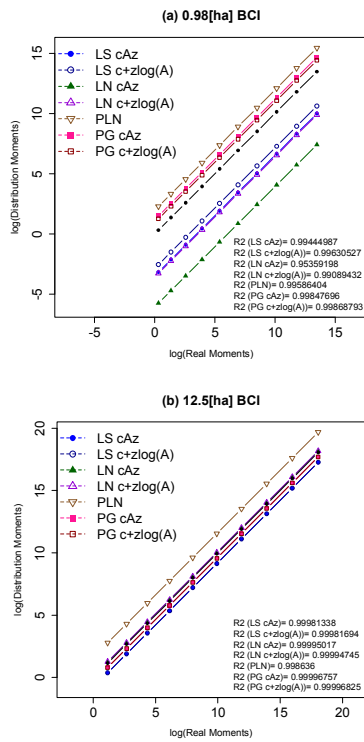
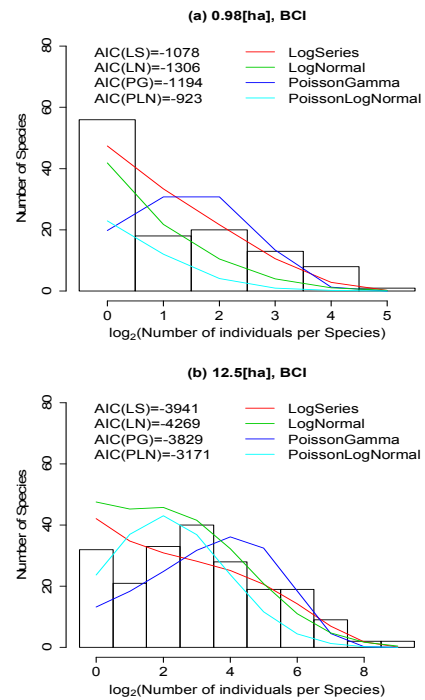


Fig. 8. Comparing linear regression of distribution moments of BCI and the real moments of order 1 up to order 10 in log-double log scale by considering area sizes 0.98, 12.5, 25.205, and 50 ha respectively for (a)-(d). In each plot, the closer width from the origin is to zero the better fitting occurs.

### B. Fitting Distributions

In Fig. 9 we show the fitting of the several distributions to the SAD for different area sizes. Observe that for larger areas the fitting provided by the lognormal is clearly the best.



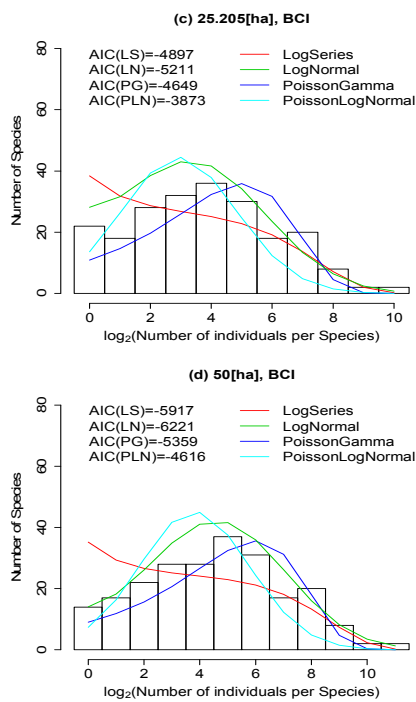


Fig. 9. Comparing the species abundance distribution of BCI, histogram, and the distributions for the area sizes of 0.98, 12.5, 25.205, and 50 ha, respectively (a)-(d).

## V. CONCLUSIONS

None of the distributions considered gave the best fit at all areas. For small areas the distribution that gave the best fit was the Poisson gamma, and for larger areas the lognormal provided the best fit. Surprisingly, the logseries did not give the best fit for small areas. Notice, however, that the quality of the fit of the logseries kept increasing when area increased, suggesting that this distribution may end up providing a good fit for very large areas.

In this work we considered a linear relationship between the log of the moments and the log of area size. We could show that the better result would have been obtained if, instead, we had considered a linear relationship between log of moments and double log of area size.

## ACKNOWLEDGMENT

We thank our previous collaborators for work on related subjects.

## REFERENCES

- [1] Fisher, R.A., Corbet, A.S. and Williams, C.B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, pp.42-58.
- [2] Preston, F.W., 1948. The commonness, and rarity, of species. *Ecology*, 29(3), pp.254-283.
- [3] Bulmer, M.G., 1974. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, pp.101-110.
- [4] Engen, S. and Lande, R., 1996. Population dynamic models generating species abundance distributions of the gamma type. *Journal of Theoretical Biology*, 178(3), pp.325-331.
- [5] May, R.M., 1975. Patterns of species abundance and diversity. *Ecology and evolution of communities*, pp.81-120.
- [6] Hubbell, S.P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.
- [7] Liao, S.X. and Pawlak, M., 1996. On image analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3), pp.254-266.
- [8] Teague, M. R. 1980. Image analysis via the general theory of moments. *J. Optical Soc. Amer.*, 70(8), pp. 920-930.
- [9] Mukundan, R., Ong, S.H. and Lee, P.A., 2001. Image analysis by Tchebichef moments. *IEEE Transactions on image Processing*, 10(9), pp.1357-1364.
- [10] Borda-de-Água, L., Borges, P.A., Hubbell, S.P. and Pereira, H.M., 2012. Spatial scaling of species abundance distributions. *Ecography*, 35(6), pp.549-556.
- [11] Borda-de-Água, L., Whittaker, R.J., Cardoso, P., Rigal, F., Santos, A.M., Amorim, I.R., Parmakelis, A., Triantis, K.A., Pereira, H.M. and Borges, P.A., 2017. Dispersal ability determines the scaling properties of species abundance distributions: a case study using arthropods from the Azores. *Scientific Reports*, 7(1), p.3899.
- [12] Feller, W. 2008. *An introduction to probability theory and its applications* (Vol. 1). John Wiley & Sons.
- [13] Pielou, E.C., 1977. *Mathematical ecology*. Wiley.
- [14] Green, J.L. and Plotkin, J.B., 2007. A statistical theory for sampling species abundances. *Ecology letters*, 10(11), pp.1037-1045.
- [15] Rosenzweig, M.L., 1995. *Species diversity in space and time*. Cambridge University Press.
- [16] Johnson, N.L., Kemp, A.W. and Kotz, S., 2005. *Univariate discrete distributions* (Vol. 444). John Wiley & Sons.
- [17] Asmussen, S., Jensen, J.L. and Rojas-Nandayapa, L., 2016. On the Laplace transform of the lognormal distribution. *Methodology and Computing in Applied Probability*, 18(2), pp.441-458.
- [18] Karlis, D. and Xekalaki, E., 2005. Mixed poisson distributions. *International Statistical Review*, 73(1), pp.35-58.
- [19] Nielsen, F. and Garcia, V., 2009. Statistical exponential families: A digest with flash cards. *arXiv preprint arXiv:0911.4863*.