ELSEVIER

Journal of Experimental Marine Biology and Ecology 330 (2006) 19 – 26

Journal of EXPERIMENTAL MARINE BIOLOGY AND ECOLOGY

www.elsevier.com/locate/jembe

Are there differences in structure between marine and terrestrial assemblages?

John S. Gray [a,*], Anders Bjørgesæter [a], Karl I. Ugland [a], Kenneth Frank [b]

[a] *Marine Biodiversity Research Group, Department of Biology, University of Oslo, P.b. 1066 Blindern, 0316 Oslo, Norway*
[b] *Department of Fisheries and Oceans, Ocean Sciences Division, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, Nova Scotia, Canada B2Y 4A2*

Received 23 June 2005; received in revised form 25 July 2005; accepted 15 December 2005

Abstract

Recently, interest in species abundance (SAD) distributions has been revived by introduction of a new model, the zero-sum multinomial (ZSM). Yet detailed statistical analyses show that the model does not differ from the lognormal distribution proposed in the 1940s. These analyses were based on data from tropical trees where all individuals in a defined area were identified to species. For many ecological data sets it is not possible to identify and count all individuals in a given area. Here we compare data on marine benthos and fish assemblages with data on terrestrial microfauna and ants. We show that these assemblages show similar SAD patterns and that the SADs are best described by a two-group lognormal model. Whereas the 2-group model fitted all data sets the single group model fitted all except the tropical rainforest ants. However, tests comparing the fits to the 2-group versus the single lognormal model showed that the 2-group model was a significantly better fit to the fish and insect data. The two groups are of rare and common species and the rare group dominates in all four data sets. We suggest that the reason for this is that rare species are continuously immigrating from outside the sampled area. Data on tropical tree assemblages where complete accounts were made do not show such high dominance of rare species where the sampled area is large. We conclude that SAD patterns are similar in marine and terrestrial systems that are open to immigration and that the lognormal distribution is still a valid model for SADs.
© 2006 Published by Elsevier B.V.

Keywords: Ants from tropical forests; Insects of heathland; Marine benthos and fish assemblages; Rare species; Species abundance distributions

1. Introduction

Hubbell's (2001) book "The unified theory of biodiversity and biogeography" has revived interest in studies of species abundance distributions, (SADs). Hubbell developed a new model for SADs based on zero-sum dynamics of species and on ecological equivalence among individuals of species within assemblages, called the zero-sum multinomial distribution (ZSM). This model has created much interest since it is based on accepted biological principles rather than being merely a statistical description of SADs. However, the model has also generated controversy since it is difficult to apply, (Alonso and McKane, 2004) and it has not yet been shown to be a convincingly better fit than other models to ecological data. For example analyses of the same data set used as one of the cornerstones in the generation of the ZSM, data on tropical trees from Barro Colorado Island, Panama (BCI), (Condit et al., 2002) led McGill (2002) and Volkov et al. (2003) to the conclusion that a lognormal distribution fitted the data as well as the ZSM,

* Corresponding author. Fax: +47 22854438.
E-mail address: j.s.gray@bio.uio.no (J.S. Gray).

0022-0981/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.jembe.2005.12.014

(although Volkov et al. argued that small differences in the chi-square goodness-of-fit test were in favour of the ZSM).The lognormal distribution is an old pattern first suggested by Preston (1948) and since then has been shown to fit very many data sets (see May, 1981 and Magurran, 2004 for reviews). Although it can be derived from biological arguments (Sugihara, 1980; Tokeshi, 1990) there are powerful arguments that the lognormal SAD pattern is merely a statistical descriptor (May, 1981; Pielou, 1975). These arguments are based on the Central Limit Theorem which states that the product of a variety of interacting factors is a lognormal distribution. May (1981) cites the GDP of countries as a good example.

In the recent spate of testing models and statistical fits to SADs most attention has been given to remarkably few data sets such as the comprehensive data available on tropical tree species richness, at BCI and Pasoh, Malaysia (McGill, 2002; Condit et al., 2002; He and Legendre, 2002) and data on breeding birds in UK (Nee et al., 1991; Gregory, 1994, 2000) and North America (Robbins et al., 1986; McGill, 2002). One of the characteristics of such data is that they are actual counts and not samples from an assemblage of unknown species richness. Recently we, (Gray et al., 2005) have shown that there are differences in SAD structure between terrestrial, (data on tropical trees at BCI) and marine species (benthos from the continental shelf of Norway) as shown by the relative importance of rare compared with common species. We suggested that a possible cause of such differences is that in marine benthos one is sampling a portion of the total species pool and thus rare species dominate at all spatial scales studied.

It is difficult to obtain raw data sets of complete assemblages since they are only rarely published in full. We have analyzed many data sets from undisturbed marine benthos assemblages and the SAD patterns are remarkably consistent. The benthic assemblages cover a wide variety of taxa. Here we compare the SAD pattern for a new data set from a benthic assemblage with that of a single taxon, marine fish (both elasmobranchs and teleosts). The tropical tree data from BCI and Pasoh have been extensively analysed so we wished to obtain data from different terrestrial assemblages. The terrestrial data sets examined here are from all microfauna of a *Calluna* heath habitat, (Delany, 1956) and are thus approximately similar to the marine benthic assemblage and from a single taxon, ants sampled on forest floors in the tropics (Longino et al., 2002), broadly comparable to the marine fish. In both the marine and terrestrial environments the samples analyzed were taken from a large regional species pool with large spatial extent. Thus these data sets are expected to show similar SADs.

We appreciate that this is a small number of data sets with which to compare marine and terrestrial data and many more are needed before conclusive statements can be reached. Yet the results are similar to our analyses of the BCI data (Gray et al., 2005) and convince us that the observed patterns may be general.

2. Materials and methods

The data sets analyzed in this paper are:

2.1. Benthos from the Hanz region of the Norwegian continental shelf

The data were sampled as part of routine monitoring of the effects of the oil industry on the marine environment. It was a baseline survey and the sites are unaffected by oil activities. The data are from a survey done in 1996. The depth ranged from 115–120 m and the sediment was fine sand with a silt-clay content varying from 5–11%; thus the habitat is rather uniform. Five 0.1 m^2 samples were taken at each of 18 sites within an 8×8 km area. The fauna retained on a sieve with 1 mm pores was identified to species where possible.

2.2. Fish from the Scotian Shelf, (Shackell and Frank, 2003)

A fishery-independent ground-fish trawl survey of the Scotian Shelf and Bay of Fundy has been conducted annually during July since 1970 by the Canadian Department of Fisheries and Oceans (coincides with Northwest Atlantic Fisheries Organization statistical divisions 4VWX). The survey follows a standardized sampling protocol using a Western IIA otter trawl with a 19-mm cod-end mesh lining towed for 30 min over a distance of approximately 3.25 km. The area swept by the trawl during this sampling interval was 0.0404 km^2. Tow locations are assigned according to a random sampling design with areas stratified by depth zones. Sampling was conducted around the clock. Because there was no fixed schedule, the allocation of day and night samples varied randomly among strata and years. In general, four weeks were required to complete a single survey.

Catch rates are standardized by tow distance and expressed as average number per tow per stratum or strata groupings. The complete survey comprises 5526 tows, with an average of 173 tows per year (range 110–213). Here the data analyzed are the 2002 survey comprising

Table 1
Overview of the data analyzed

	Marine benthos	Fish Scotian shelf	Heathland invertebrates	Tropical rainforest ants
No. of species	291	79	76	197
No. of individuals	24,993	149,680	717	1,415
No. of samples	18	208	7	41
Study area (km^2)	25	170,000	?	?
Area sampled	9 m^2	104 h	85 m^2	246 l

208 tows since this was the newest data available and we wished to analyze data for a single year.

2.3. Invertebrates from pioneer heath in UK (Delany, 1956)

The data are collected from pioneer heath (*Calluna vulgaris* Salisb.) at three localities in southern England in Devonshire, Hampshire and on the Island of Lundy, a distance of ca. 200 km. The samples were taken by randomly casting quadrats 0.5 m^2 with a total area covered of 12.5 m^2 at each sites (save site 7 which only covered 10 m^2). Thus the total area sampled was 85 m^2. The fauna <ca. 5 mm in size was collected carefully by hand.

2.4. Ants from tropical rain forest (Longino et al., 2002)

Ants (Hymenoptera: Formicidae) were sampled in a tropical rain forest in the La Selva biological Station, Heredia Province, Costa Rica as part of the Arthropods of La Selva (ALAS) project (http://viceroy.eeb.uconn.edu/ALAS/ALAS.html), (Longino et al., 2002). Details of the sample area can be found on the website. The data on soil and litter fauna were collected using eight different methods. Here we use data from samples taken over a 14 year period from 41 sites covering the 1500 ha of the La Selva reserve but from only one of the sampling methods, the Winkler method, since this method obtained the highest number of species. In this

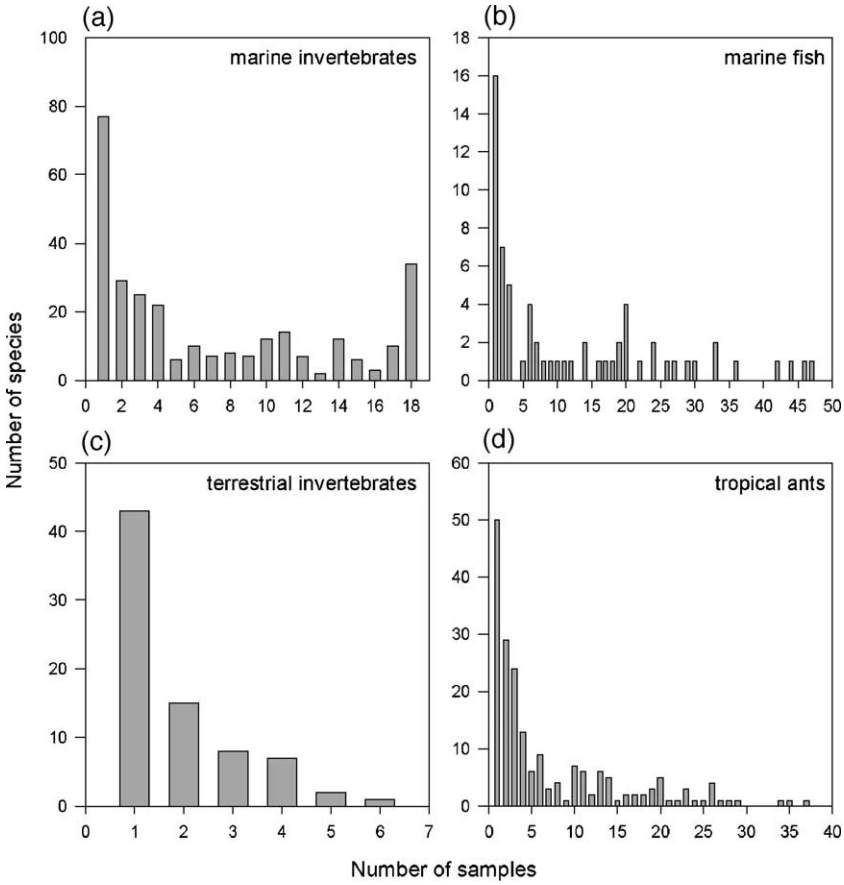

Fig. 1. Frequency distribution of species among sites occupied for a) benthos from the Hanz oilfield, Norwegian continental shelf, b) fish from the Scotian Shelf c) insects from heathland plants and d) ants from a tropical rain forest (Longino et al., 2002).

method 6l samples of sifted litter were placed in Winkler bags, which are large cotton bags with a jar of alcohol attached underneath. As the litter dries and decays over 3 days animals are trapped in the alcohol.

2.5. Plotting of SADs and determining the breakpoint

We plotted the species occurrences at sites for the data sets to illustrate the proportion of rare species present. Then we plotted species abundance distributions as the number of species against numbers of individuals per species in geometric classes (x2 classes called octaves). We used Williams (1964) modification of Preston's original method (Preston, 1948, 1962), which was also used by McGill (2002) and Volkov et al. (2003). In this method octave 1 = half the number of species occurring with 1 individual per species. Octave 2 = half the number of species with 1 individual per species plus half the number of species with 2 individuals per species. Octave 3 = half the number of species with 2 individuals per species all those with 3 individuals per species and half those with 4 individuals per species. Octave 4 = half the number of species with 4 individuals per species, all those with 5, 6 and 7 individuals per species and half of those with exactly 8 individuals per species, etc.

Plotting the data in this manner suggested that there were two (or more) lognormally distributed groups of species present rather than one single group. We developed a method to determine break-points which could separate the data into two or more groups. In order to partition the species groups objectively we first plot the number of species observed in 1, 2,…, n samples where n = total number of samples. Then we fit two regression lines to this plot starting with 2 samples Let us assume that there are 100 samples. For $k = 2, 3, 4, 5, \ldots, 99$ we then fit a straight line by linear regression to both the data set from 1 to k samples, and from $k+1$ to 100 samples. The total squared deviance is the sum of the residuals of the two regression lines and denoted as SS(k). The breakpoint k^* is determined as that number of samples that minimizes the squared deviance

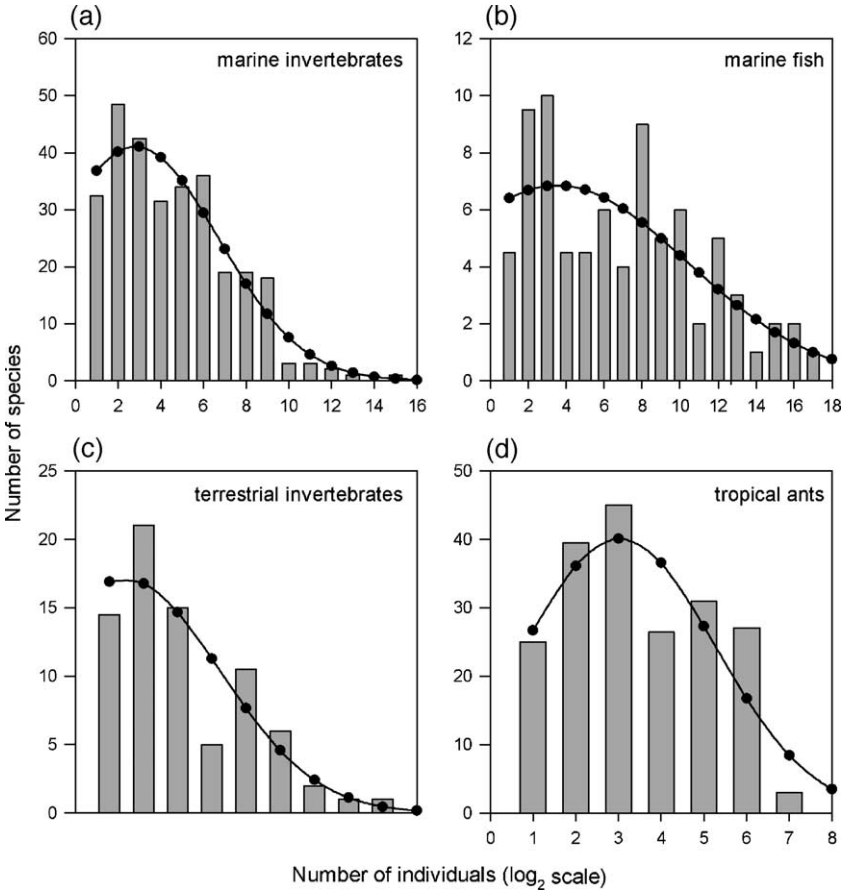

Fig. 2. SADs fitted to the same data as in Fig. 1 using Williams (1964) modification of Preston's (1948) method for calculating the SAD groupings (see Materials and methods for a detailed explanation). Curves show a single lognormal distribution fitted to these data.

Table 2

a) Test of significance of single lognormal and two-group lognormal models using χ^2 test for different data sets. Significant differences are shown in bold

Model tested	Benthic fauna		Fish Scotian shelf		Heathland invertebrates		Tropical rainforest ants	
	Single	2-group	Single	2-group	Single	2-group	Single	2-group
df	12	10	14	12	6	4	4	2
P	0.36	0.12	0.68	0.93	0.37	0.81	**0.00**	0.23
Critical-value (p=0.05)	21	18	24	21	13	9	9	6

b) Comparison of two-group versus single lognormal model using *F* ratio test. Significant differences where the two-group model is superior are shown in bold

	Benthic fauna	Fish Scotian shelf	Heathland invertebrates	Tropical rainforest ants
Ratio *F*	0.47	4.33	8.12	[a]
P	0.64	**0.04**	**0.04**	[a]

[a] Test not valid since the single group model does not fit.

(minimum deviance); i.e. SS(k*)<SS(k) for all k. The Piecewise Regression Procedure in S-Plus (Crawley, 2004) was used to derive the minimum de-viance to define the breakpoint between two linear re-gression lines fitted to the data. It is important to note that the break-point is based on the species-site patterns and not on the SAD patterns. This provides an additional level of objectivity in deciding where the break-points occur.

2.6. Fitting of lognormal models

Lognormal models were fitted to the SADs using the function NORMDIST in Excel. The "solver" procedure in Excel was then used to find the best fit. Solver minimizes the sum of the squares (SSE) by changing the mean and standard deviation of the normal curve. The settings used were 10,000 iterations, precision 0.00000001, tolerance 5% and convergence 0.0001.

For testing goodness-of-fit we used the Kolmogorov–Smirnov test and the chi^2 test. The Kolmogorov–Smirnov test showed that none of the models were significantly different from the observed data using this test, (the results are not shown). Thus the K–S test is insensitive. When using the chi^2 however, data for octaves with low expected abundances may give large values of chi^2 which lead to conclusions of lack of fit to the model. Thus we did not include values for octaves with expected values less than 1.

3. Results

Table 1 shows a summary of the data analyzed. For the marine data the samples were taken within prescribed areas and so there is an estimate of the total area that has been covered. This is not so for the terrestrial data which for the heathland invertebrates covers three sites separated by ca 200 km. The data are however, similar in a key aspect that they are not counts of all species but samples from given areas. We are interested in whether terrestrial data collected in a similar manner to the marine data showed similar SAD patterns.

Fig. 1 shows the species among samples distributions. All the assemblages are dominated by rare species, although the heath invertebrate fauna covers relatively few sites. The Hanz data is unusual for marine benthos since there are a high number of species (38) occurring at 18 samples. We have no explanation for this and so many species occurring at a high number of samples is not a typical pattern for marine benthic data (see Ellingsen and Gray, 2002; Gray et al., 2005 for more typical data). Yet the pattern for occurrence for the rare and moderately common species is typical for benthic assemblages. Species Abundance Distributions (SADs) were then calculated for these data and a single lognormal was fitted (Fig. 2).

Fig. 2 shows that for all the assemblages although a curve can be fitted there are clear divergences from a single lognormal model as the data are "bumpy" suggesting that more than one lognormal underlies the patterns. The lack of fit of the single group model is shown in Table 2.

Two-group lognormal models were then fitted to these data and are shown in Fig. 3.[1] Table 2a) shows the

[1] The invertebrate fauna collected from heathland covered too few classes (Fig. 1) to use the species-site occurrences as a basis for determining the break-point. Instead the break-point was fitted by eye, based on the SAD-pattern (or by the method described above on the SAD-pattern, but in this case excluding the first octave).

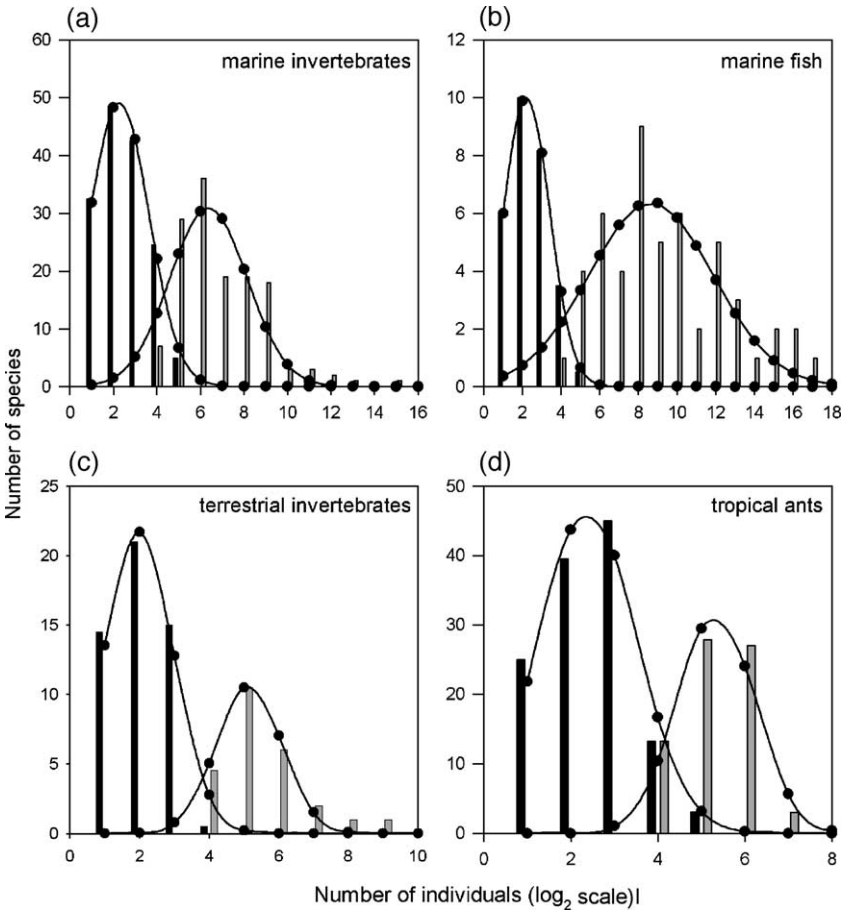

Fig. 3. The fits of the two-group model compared to observed species abundance distributions for groups of rare and common species, data sets as in Fig. 1.

goodness-of-fit tests. The 2-group lognormal model fits all data, whereas the single lognormal fits all the data save tropical rainforest ants. We then tested the hypothesis that the 2-group model was a better fit then the single lognormal. For two data sets, fish from the Scotian shelf and heathland invertebrates, the 2-group model was a significantly better fit. Both models fitted the benthic data equally well whereas for tropical forest ants it had already been shown (Table 2a) that the single lognormal model did not fit and thus the 2-group model is the best fit. For three of the four data sets the 2-group model was a significantly better fit whereas with the benthic data both the single and 2-group models fitted.

Fig. 3 shows the 2-group model fitted to the data. Both the marine and terrestrial data sets are remarkably similar, with a dominance of the rare group of species in the first few octaves. Clearly the same structuring forces dominate in all of the assemblages studied.

4. Discussion

Much effort has been concentrated on analyzing SAD patterns for a few data sets where all the species have been counted, notably the BCI and Pasoh forest tropical tree species data. There is dispute as to whether these data sets fit a simple single lognormal distribution or the ZSM (Hubbell, 2001; McGill, 2002; Volkov et al., 2003). In a recent paper (Gray et al., 2005) we analyzed data from samples rather than on exact counts of all species and found that the BCI data and data on marine benthos from the Oseberg oil field Norway fitted a 2- or 3-group lognormal model where the groups were rare and common species or rare, intermediate and common in the 3-group model.

The analyses done here show that for the studied assemblages, the data clearly do not fit a single lognormal distribution (Figs. 2 and 3, Table 2). Williamson and Gaston (2005) have argued that a single lognormal

distribution is not an appropriate fit to data from most natural assemblages. Yet they did not test the two-group lognormal model. Earlier Ugland and Gray (1982) argued for a 3-group model and Magurran and Henderson (2003) showed that fish assemblages taken over time in the Severn estuary showed a clear two-group pattern, although they argued for a log-series distribution for the rare species and a lognormal for the common species. The occurrence of mixtures of lognormal groups has been well-studied in other fields such as meteorology (Kayano and Shimizu, 1994) and medicine (distribution of red-blood cell sizes) McLaren et al., 1991) and the statistical methods to detect underlying groups were developed some time ago (Fowlkes, 1977).

The occurrence of a 2-group model is not confined to data where only samples were taken and not complete counts. Gray et al. (2005) reanalyzed the BCI data at different spatial scales and showed that in addition to a single lognormal a 2-group model fitted the data at scales from 5, 10, 25 and 50 ha, with the dominance of the rare group of species being greater at the smallest scales. We believe that this is due to immigration of rare species from the species pool outside the area sampled. In assemblages that are only sampled (as opposed to counts where all individuals are identified and counted) the rare group of species dominates since there is constant immigration of species from the species pool outside the area sampled.

The lognormal distribution fits an extremely wide range of natural phenomena (Limpert et al., 2001) ranging from concentrations of heavy metals in the environment, to latency periods for diseases, rainfall patterns and fungicide sensitivity. Limpert et al. (2001) show that the lognormal results from the multiplicative version of the central limit theorem which proves that "the product of many independent, identically distributed, positive random variables has approximately a lognormal distribution." In addition they note that the relevant operation for combining data to obtain a lognormal is multiplication and "the product of independent lognormal quantities also follows a lognormal." For these reasons we believe that Williamson and Gaston's (2005) pronouncement of the "death" of the lognormal is premature!

Our arguments for the expectation of lognormal SADs are based on the assumption that the population of a species and its dispersal are affected by a wide variety of physical, chemical and biological factors which act in a multiplicative manner giving a lognormal distribution. In an assemblage of many species however, the underlying pattern may be determined by two or more groups of species, which are more or less "fused" into a single lognormal distribution. To what extent the groups are independent remains to be studied in detail. Our initial analyses (Gray et al., 2005) suggest that for the BCI data the groups are independent whereas for the marine benthos data as larger samples are taken the formerly rare species are found among all the octave classes. This suggests that in tropical tree assemblages rare immigrant species have difficulty in establishing large population sizes whereas in assemblages more open to immigrant species, rare immigrants are more easily able to establish large population sizes.

Finally, it must be acknowledged that analyses based on only a few data sets may not be representative. However, we have shown (Gray et al., 2005) that the BCI tropical tree data and a marine benthic assemblage fit the two-group lognormal model. The analyses done here on specific taxa (fish and ants) and on assemblages (benthos and heathland microfauna) also show fits to the two-group lognormal model. Whilst our analyses cover only a small number of assemblages and taxa we hope that this study may stimulate others to investigate other data sets and taxa to test the generality of the findings.

Acknowledgements

We thank the Norwegian Oil producers National Organization (OLF) for allowing us to use data from their environmental database and to John Longino for allowing us to use his data on tropical rainforest ants. This research was supported with funds from the University of Oslo. The paper is a contribution to the European Union Framework 6 Network of Excellence in Marine Biodiversity (MarBEF). **[SS]**

References

Alonso, D., McKane, A.J., 2004. Sampling Hubbell's neutral theory of biodiversity. Ecol. Lett. 7, 901–910.

Condit, R., Pitman, N., Leigh, J.E.G., Chave, J., Terbrough, J., Foster, R.B., et al., 2002. Beta-diversity in tropical forest trees. Science 295, 666–669.

Crawley, M., 2004. 2002 Statistical Compuing: An Introduction to Data Analysis Using S-Plus. Wiley, Chchester. UK.

Delany, M.J., 1956. The animal communities of three areas of pioneer heath in southwest England. J. Anim. Ecol. 25, 112–126.

Ellingsen, K.E., Gray, J.S., 2002. Spatial patterns of benthic diversity: is there a latitudinal gradient along the Norwegian continental shelf? J. Anim. Ecol. 7, 373–389.

Fowlkes, E.B., 1977. Unpublished Bell Telephone Laboratories Technical Memorandum.

Gray, J.S., Bjørgesæter, A., Ugland, K.I., 2005. The impact of rare species on natural assemblages. J. Anim. Ecol. 74, 1131–1139.

Gregory, R.D., 1994. Species abundance patterns of British Birds. Proc. Roy. Soc. Lond. Ser. B. Biol. Sci. 257, 299–301.

Gregory, R.D., 2000. Abundance patterns of European breeding birds. Ecography 23, 201–208.

He, F., Legendre, P., 2002. Species diversity patterns derived from species-area models. Ecology 83, 1185–1198.

Hubbell, S.P., 2001. The Unified Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.

Kayano, K., Shimizu, K., 1994. Optimal thresholds for a mixture of lognormal distributions as the continuous part of the mixed distribution. J. Appl. Meteorol. 33 (12), 1543–1550.

Limpert, E., Stahel, W., Abbt, M., 2001. Lognormal distributions across the sciences: keys and clues. Bioscience 51, 341–352.

Longino, J.T., Coddington, J., Colwekll, R.K., 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. Ecology 83, 689–702.

Magurran, A.E., 2004. Measuring Ecological Diversity. Blackwell, Oxford.

Magurran, A.E., Henderson, P.A., 2003. Explaining the excess of rare species in natural species abundance distributions. Nature 422, 714–716.

May, R.M., 1981. Patterns in multi-species communities. In: May, R.M. (Ed.), Theoretical Ecology: Principles and Applications. Blackwell, Oxford, pp. 197–227.

McGill, B.J., 2002. A test of the unified neutral theory of biodiversity. Nature 422, 881–885.

McLaren, C.E., Wagstaff, M., Brittenham, G.M., Jacobs, A., 1991. Detection of two-component mixtures of lognormal distributions in grouped, doubly truncated data: analysis of red blood cell volume distributions. Biometrics. 47 (2), 607–622, Jun.

Nee, S., Harvey, P.H., May, R.M., 1991. Lifting the veil on abundance patterns. Proc. Roy. Soc. Lond. Ser. B. Biol. Sci. 243, 161–163.

Pielou, E.C., 1975. Ecological Diversity. John Wiley and Sons, New York, USA.

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

Preston, F.W., 1962. The canonical distribution of commonness and rarity, of species. Ecology 29, 254–283.

Robbins, C.S., Bystrack, D., Geissler, P.H., 1986. The Breeding Bird Survey: Its First Fifteen Years, 1965–1979. US Department of the Interior Fish and Wildlife Service, Washington, DC.

Shackell, N.L., Frank, K.T., 2003. Marine fish diversity on the Scotian Shelf Canada. Aquat. Conserv: Mar. Freshw. Ecosys. 13, 305–321.

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

Tokeshi, M., 1990. Niche apportionment or random assortment: species abundance patterns revisited. J. Anim. Ecol. 59, 1129–1146.

Ugland, K.I., Gray, J.S., 1982. Lognormal distributions and the concept of community equilibrium. Oikos 39, 171–179.

Volkov, I., Bananver, J.R., Hubbell, S.P., Maritan, A., 2003. Neutral theory and relative species abundance in ecology. Nature 424, 1035–1037.

Williams, C.B., 1964. Patterns in the Balance of Nature and Related Problems in Quantitative Ecology. Academic Press, London.

Williamson, M., Gaston, K.D., 2005. The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. J. Anim. Ecol. 74, 409–422.