

Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model

Karl I. Ugland,¹ P. John D. Lamshead,² Brian McGill,³
John S. Gray,¹ Niall O'Dea,⁴ Richard J. Ladle⁴
and Robert J. Whittaker^{4*}

¹*Department of Biology, University of Oslo, Oslo, Norway,* ²*Department of Zoology, The Natural History Museum, London, UK,* ³*Department of Biology, McGill University, Montreal, Quebec, Canada and* ⁴*Biodiversity Research Group, Oxford University Centre for the Environment, Oxford, UK*

ABSTRACT

Problem: One of the longest standing problems in ecology has been the identification of an ecologically meaningful model that adequately and parsimoniously describes the relationship between the number of species and the numbers of individuals of each species in a sample.

Methods and results: Here, we develop a new species abundance distribution model based on the gamma distribution and termed the Gambin model. In this model, a single variable, α , determines the shape of the gamma distribution and the 'dimensionality' of the sampled community. Using 33 existing data sets, we show that the model produces species abundance distributions that can be fitted to a wide range of marine and terrestrial data and that the Gambin model outperforms three prominent alternative models: the log-series, lognormal, and zero sum multinomial models.

Conclusions: The Gambin α parameter provides a form of diversity metric, reflecting the complexity of a community's interactions with its environment.

Keywords: Gambin model, Gamma distribution, lognormal distribution, log-series distribution, macroecology, species abundance distribution.

INTRODUCTION

It has long been recognized that the distributions of species abundances within ecological communities – or metacommunities (*sensu* Magurran, 2005) – show consistent patterns (Motomura, 1932; Fisher *et al.*, 1943), reflecting emergent (macroecological) outcomes of the highly complex interactions between the constituent organisms and their environment. Thus, species

* Address all correspondence to R.J. Whittaker, Biodiversity Research Group, Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY, UK. e-mail: robert.whittaker@ouce.ox.ac.uk
Consult the copyright statement on the inside front cover for non-commercial copying policies.

abundance distributions (SADs) from natural or unstressed communities generally approximate a lognormal distribution (Preston, 1948, 1962; Williams, 1964), while communities that are relatively new or have recently been ecologically heavily ‘stressed’ often produce SADs that conform closer to a log-series distribution (May, 1975; Whittaker, 1975). This latter situation is driven by a few species becoming relatively more dominant while many species become proportionally rare. Although the lognormal and log-series models have been seen as purely stochastic fits to species abundance distributions, and hence difficult to interpret biologically (Magurran, 2004), they have nevertheless proved useful and are still widely used to describe and explore the properties of communities (Magurran and Henderson, 2003).

In many ways, Hutchinson’s (1957) ecological niche theory represents the converse situation, in which an elegant biological concept has proved difficult to apply in practice. The niche was defined by Hutchinson as the sum of all the environmental factors acting on the organism; the niche so defined is a region of an n -dimensional hyperspace (Hutchinson, 1957). The niche concept has been remarkably influential in the development of ecological thought, but is difficult to apply in practice because species are affected by a bewildering number of processes, or dimensional axes of the environment, and are constrained within their fundamental niches to more limited or ‘realized’ niche space by biotic interactions. When considering the relationships of an assemblage of locally interacting species to environment, niche theory further holds that each species should occupy a slightly different space in the n -dimensional hypervolume of the community. Herein, we consider how the dimensionality of the community may be reflected in the form of the SAD of the sample, and we present a new SAD model, termed the Gambin model (derivation explained below), which we suggest provides a simple means of quantifying the ‘reach’ of the constituent species in the community into the hypervolume.

The Gambin (Gamma-binomial) model is a stochastic model based on the Gamma distribution. The model involves a single parameter, α , describing the shape of the species distribution curve and which is assumed to represent the ‘dimensionality’ of the sampled community. In addition to providing a formal description of this model, we also compare the performance of this new model to the two ‘industry standard’ models and one other prominent SAD model, Hubbell’s (2001) zero-sum multinomial (ZSM).

It should be noted that when a stochastic model is fitted to data, an extra variable or loss of degree of freedom is required to scale the model to the data. So although having only one parameter, Gambin technically becomes a two-parameter model when fitted to real SADs. This applies to all models of this type. For simplicity, we follow the convention of referring to the log-series and Gambin models as one-parameter models, and the lognormal as a two-parameter model, even when these models are fitted to data.

Our comparison of the performance of Gambin, log-series, and lognormal models is based on 33 data sets culled from the literature and from our own previous work. We also compare the Gambin model with Hubbell’s (2001) ZSM (‘neutral’) model for eight data sets. The ZSM is a three-variable model – ‘J’, ‘m’, and either ‘theta’ or ‘S’ depending on the fitting method (Hubbell, 2001; McGill *et al.*, 2006). We would not normally choose to compare a one-variable and three-variable model in this way, as the latter should easily outperform the former when fitted to data, but the ZSM is a recent development that has aroused general interest (McGill, 2003; Chave, 2004; McGill *et al.*, 2006) and hence merits attention herein. Note that when fitted, the ZSM also acquires an extra, (fourth) scaling parameter as discussed above.

METHODS

General description

The relative abundance of species in an artificial community of a given richness can be modelled with the binomial distribution. If the abundance is measured on a logarithmic scale, each species is placed in an octave determined by the number of successes in a Bernoulli series with probability P and a number of trials equal to the maximum octave in the sample. However, the binomial probability P need not be the same for all species, so we assume that it has a Gamma distribution. As the Gamma distribution has no end-point, we impose an arbitrary cut-off on the x axis such that the area under the curve over the interval from zero to the cut-off point is 0.99 (i.e. the cut-off point is the 99% quantile). The interval between 0 and the cut-off point is then partitioned into 100 sections to smooth the final curve. A low relative success on the axis leads to a low expected species abundance (i.e. a low P -value), whereas a high relative success leads to a high expected species abundance (i.e. a high P -value). This is quantified by the binomial distribution, hence the name Gambin (Gamma-binomial). A random value of the Gamma-variate will place the species in one of the 100 sections, for example section 28. The corresponding binomial value is then set to the section number divided by 100 (for example, if the 28th section was selected, the binomial probability is set to $P = 0.28$).

Although the Gamma distribution has two parameters, it should be emphasized that the Gambin model has only one when the maximum octave is fixed. Intuitively, this is because the concept of success is a relative measure – that is, the relative abundances are independent of the counting unit. The Gamma distribution becomes scale-independent when it is cut off at the 99% point – that is, all values of the scale parameter give the same distribution over the 100 intervals. Hence, we use scale parameter $\beta = 1$ and are left with only one parameter, the curve shape α . High α -values correspond to a community where most species are relatively abundant and low α -values to a community where most species are relatively rare (see Fig. 1).

Formal mathematical description

Fixing the scale parameter equal to 1, the 99% point for the shape parameter α is defined as the number $c_1 = c_1(\alpha)$ that covers an area of 0.99:

$$\int_0^{c_1(\alpha)} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx = 0.99$$

The interval is subdivided between 0 and $c_1(\alpha)$ into 100 equal sub-intervals: $(k/100) * c_1$ for $k = 1, \dots, 100$. Since these sub-intervals have a natural linear ordering, $I_1 < \dots < I_{100}$, they may be taken to represent a relative order of the species' success. That is, the species falling in the interval I_1 have the lowest success, and so on up to the most successful species in interval I_{100} . Thus the frequencies of the relative successes will be arranged according to:

$$P\left\{\frac{k-1}{100} < \text{Rel. success} < \frac{k}{100}\right\} = P\{I_{k;\alpha}\} = \left[G_{\alpha,1}\left(\frac{k}{100} c_{\alpha,1}\right) - G_{\alpha,1}\left(\frac{k-1}{100} c_{\alpha,1}\right) \right] / 0.99$$

where G is the Gamma distribution with shape parameter α and scale parameter 1.

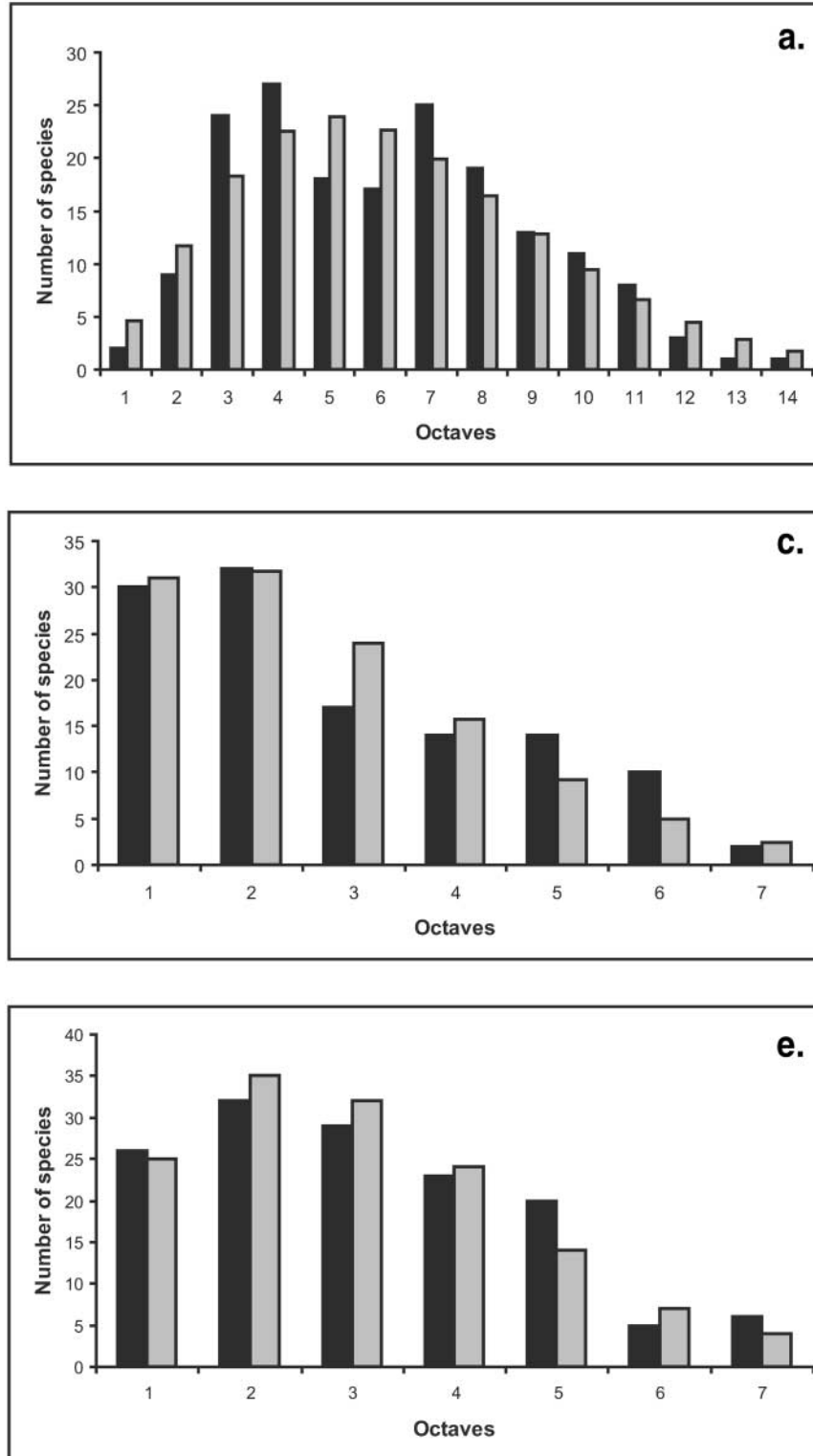
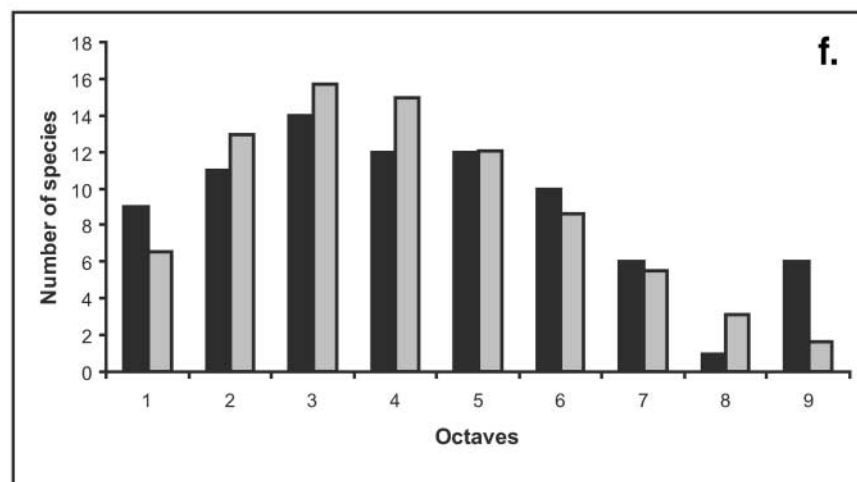
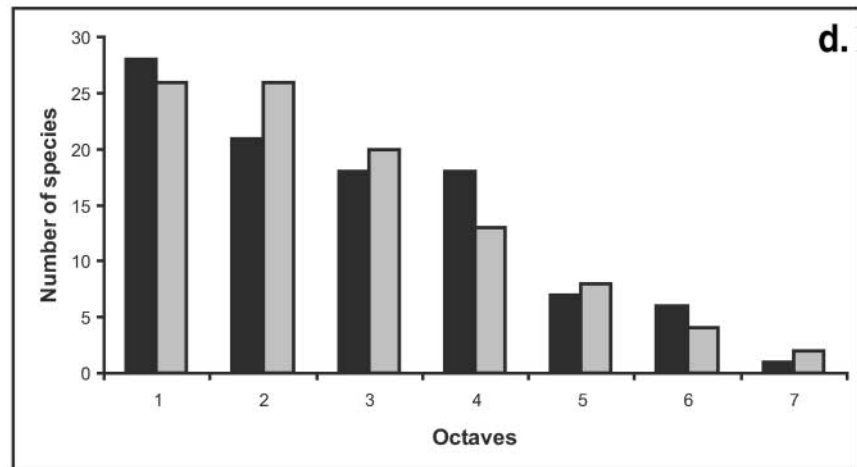
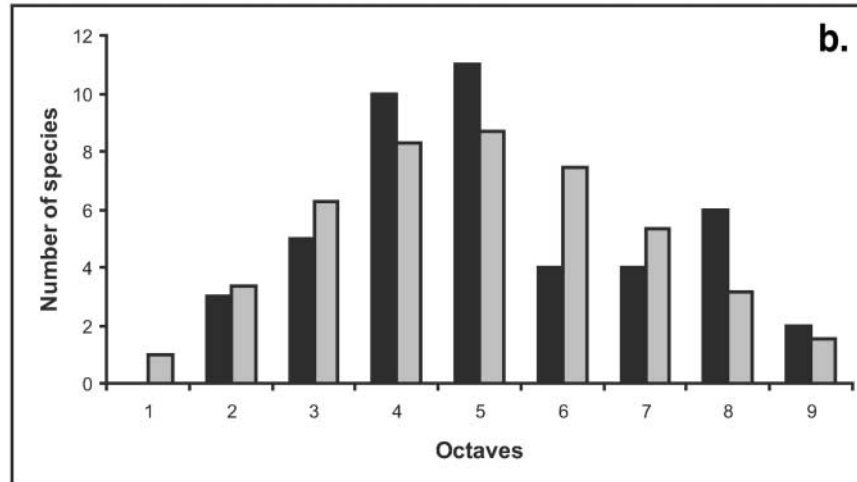


Fig. 1. Species abundance distribution curves for field data (black bars) and the Gambin model (grey bars) for six illustrative data sets: (a) Irish Sea nematodes, $\alpha = 4.0$, $P = 44\%$; (b) freshwater insects on *Callitriche*, $\alpha = 7.9$, $P = 64\%$; (c) deep sea nematodes, $\alpha = 1.8$, $P = 54\%$; (d) Ecuadorian birds in



agricultural land, $\alpha = 1.8$, $P = 62\%$; (e) Ecuadorian cloud forest birds in secondary forest, $\alpha = 2.7$, $P = 89\%$; (f) non-disturbed Loch Linnhe benthic fauna, $\alpha = 3.7$, $P = 79\%$. See Table 1 for data sources.

Translation of these quantities into frequencies of abundances requires a link function to the octaves. The simplest transformation is obtained by postulating that the abundance of all species within each of the 100 sub-intervals has a Bernoulli distribution: the species belonging to the k -th interval I_k are given a probability of success equal to $p_k = k/100$, and the number of trials is one less than the number of octaves: $n_{\text{oct}} - 1$ (the reason for subtracting 1 is that the total number of octaves is the sum of the maximum octave and the zero-th octave containing those species with no success – which on a logarithmic scale means just one individual).

Summing over the various species groups gives the following compact expression for the frequency of species in the s -th octave:

$$A_s = \sum_{k=1}^{100} P\{I_{k;u}\} \binom{n_{\text{oct}} - 1}{s} \left(\frac{k}{100}\right)^s \left(1 - \frac{k}{100}\right)^{n_{\text{oct}} - 1 - s}$$

Deviance test

We applied a simple simulation test, the deviance test, which can be used to find the best fit for each of the three models to real data sets, using Microsoft Excel solver. For the deviance test, 100 intervals are drawn randomly with the Gamma distribution having the estimated shape parameter α . Next, we draw a random integer between 0 and the maximum octave number, using the Bernoulli distribution, having a probability of success equal to the interval number divided by 100. This is done for each species and the simulated histogram is then compared with the expected histogram (obtained with the α -value used in the simulations) by calculating the deviance, defined as the sum of the squared differences between each pair of octaves in the two histograms. To calculate the error distribution, 1000 random deviances are simulated. The P -value of the dimension $\hat{\alpha}$ may then be estimated as the rank of the deviance of the observed histogram divided by 1000.

The zero-sum multinomial model

Hubbell's (2001) model presents computational difficulties. Previous methods (Volkov *et al.*, 2003) maximize fit by varying m with the constraint that $S_{\text{predicted}} = S_{\text{observed}}$ (m and S imply theta). Here, we simplify by maximizing fit by varying m and theta with no constraint on S (although theta implies S). This allows us to test the ZSM against a greater variety and quantity of data, by fitting larger data sets, by two orders of magnitude, than has been previously undertaken for the ZSM: but analysis of some very large data sets remains computationally impractical. This modification produces better fits than the first method, making our comparison between Gambin and the ZSM conservatively in favour of the ZSM.

Data sources

We searched for data sets suitable for the present analyses. In short, we sought well-specified SAD data sets from the literature. In addition, we included several data sets collected/collated by authors of the present paper, the properties of which we were thus very familiar with (these are available from the authors). The data sets include plant and animal, marine and terrestrial systems, and span a wide variety of ecological circumstances and properties (Table 1).

RESULTS

The Gambin model produces a range of species distributions matching the range of SAD curves that have been observed in nature (Fig. 1). More specifically, values approximating $\alpha = 4$ feature a near central modal curve, similar to a lognormal distribution; $\alpha = 1$ produces a strongly left-skewed curve, with the mode in the first bin; and a value of $\alpha = 2$ produces a transitional shape.

Gambin was fitted to natural SADs and its performance compared with traditional empirical models, the lognormal and the log-series, for the 33 data sets, obtained from the literature, using a deviance test (Table 1). Gambin produced curves that were not significantly different from the data for 26 (79%) of the data sets: the figure for the lognormal was 11 (33%) and that for the log-series was 17 (52%).

We also fitted the ZSM to eight selected data sets (Table 1). The eight were chosen partly to cover a range of types of assemblages and a range of Gambin α -values (from 1.2 to 6). The ZSM was not significantly different from five (63%) of the eight SADs; in contrast, Gambin was not significantly different from six (75%) of these. Given that the Gambin model is more parsimonious (fewer parameters) than the ZSM, these analyses indicate that Gambin outperforms all three of the alternative SAD models considered herein.

DISCUSSION

Our analysis suggests that the Gambin model is able to produce a range of species distributions closely approximating empirical SAD curves from 33 ecologically diverse data sets, including plants, invertebrates, and vertebrates, and from marine, freshwater, and terrestrial ecosystems (see Fig. 1). Although this is not the first 'general' model (cf. Diserud and Engen, 2000) to provide a better fit than the two normative models within ecology (lognormal and log-series), we suggest that the Gambin model is the most parsimonious general model proposed to date, and has the added advantage of generating a single parameter, α , describing the shape of the resultant species abundance curve.

It might be argued that inventing and testing a new species abundance distribution amounts to no more than a 'curve-fitting' exercise, and that the above results are therefore not persuasive of the ecological meaning of the Gambin model. We therefore offer the following theoretical interpretation.

The distribution of abundances among species within a community reflects resource use by the species individually and collectively and can be conceptualized in terms of Hutchinson's (1957) n -dimensional hypervolume. Historically, Hutchinson's dimensions have typically been associated with readily quantifiable ecological parameters (MacArthur, 1972). Resources were treated as either an aggregated quantity represented along a single one-dimensional resource-axis or as a quantity that might be subdivided in more than one dimension and represented along two or more axes (e.g. height above the ground, temperature, pH, or size of food particles). Measurements of ecological parameters have often failed to explain species distributions, perhaps because the measurements are too anthropocentric (i.e. parameters are chosen based upon incorrect supposition or the availability of sampling equipment). Moreover, there is often an implicit assumption in the literature that the parameters are independent, whereas real communities are likely to be influenced by limited numbers of 'bundles' of parameters interacting positively, negatively, linearly, and non-linearly. Systems structured by one or a few such bundles of

Table 1. Comparison of the fit of the Gambin model, lognormal, and log-series to 33 data sets

Location and type of data	No of species	GM $\alpha/P\%$	LN $P\%$	LS $P\%$	ZSM $P\%$
Terrestrial plants					
Soil seed bank of a secondary tropical rainforest (Butler and Chazdon, 1998)	34	6.0 / 69	19	2.1	36
BCI tropical trees > 10 cm (Hubbell, 2001)	225	5.0 / 37	8	0.2	99
BCI tropical trees > 1 cm (Hubbell, 2001)	923	7.2 / <0.1	<0.1	<0.1	
Pasoh tropical trees (Hubbell, 2001)	677	4.1 / 3.3	<0.1	<0.1	
Ugandan forest, primary woodland (Eggeling, 1947)	35	1.0 / 13	<0.1	14	
Ugandan forest, colonizing woodland (Eggeling, 1947)	34	1.2 / 73	20.7	30	
Ugandan forest, ecotone woodland (Eggeling, 1947)	58	1.7 / 61	<0.1	3.8	
Terrestrial vertebrates					
British birds (Gaston and Blackburn, 2000)	217	2.7 / <0.1	13	<0.1	
Ecuadorian birds in 'primary' montane forest (O'Dea <i>et al.</i> , 2006) ⁺	125	1.5 / 92	<0.1	78	
Ecuadorian birds in 'secondary' montane forest (O'Dea <i>et al.</i> , 2006) ⁺	141	2.7 / 89	<0.1	24	
Ecuadorian birds in montane forest edge habitats (O'Dea <i>et al.</i> , 2006) ⁺	132	2.1 / 61	<0.1	68	
Ecuadorian birds in montane agricultural habitats (O'Dea <i>et al.</i> , 2006) ⁺	99	1.8 / 63	<0.1	91	
Terrestrial invertebrates					
Heteroptera (Southwood, 1996)	47	5.7 / 56	35	31	
Rothampstead moths (Williams, 1964)	197	1.6 / 84	<0.1	36	
Malaysian butterflies (Williams, 1964)	620	2.6 / 0.2	<0.1	<0.1	
Freshwater invertebrates on macrophytes					
<i>Carex</i> community (Harrod, 1964)	27	3.6 / 10	9	0.2	
<i>Ranunculus</i> community (Harrod, 1964)	33	5.4 / 92	21	3	
<i>Callitriche</i> community (Harrod, 1964)	45	7.9 / 64	16	0.5	
<i>Veronica</i> community (Harrod, 1964)	30	6.0 / 39	18	1.2	

Marine vertebrates						
Nova Scotia bathyal fish data (Shackell and Franck, 2003)	146	1.2 / 11	<0.1	<0.1	<0.1	
Hinkley Point, England, estuarine fish (Magurran and Henderson, 2003)	81	1.6 / 91	3.2	43		
Marine macrofauna						
Norwegian shelf (Gray <i>et al.</i> , 1990)	2207	1.6 / <0.1	<0.1	<0.1	<0.1	
Frierfjord, Norway, all data (Gray <i>et al.</i> , 1988)	110	1.7 / 36	0.2	11		
Port Philips Bay, Australia, coastal mud (Currie and Parry, 1999)	303	1.3 / 47	<0.1	<0.1		
Port Philips Bay, Australia, coastal sand (Currie and Parry, 1999)	542	2.3 / 0.45	<0.1	0.5		
Bass Straight, Australia, coastal data (Currie and Parry, 1999)	801	2.2 / <0.1	<0.1	29	61	
Loch Linnhe, Scotland, control samples, 1963–1965 (Pearson, 1975)	81	3.7 / 79	11	30		
Loch Linnhe, Scotland, polluted samples, 1970–1972 (Pearson, 1975)	54	2.9 / 97	51	59		
Marine nematodes						
Firth of Clyde, Scotland, sandy beach, low water spring (Lamshead, 1986)	113	3.2 / 13	1.2	19		
Dublin Bay, Irish Sea, 47 m depth (Lamshead and Boucher, 2003)	178	4.0 / 44	0.6	0.1	2.5	
San Diego Trough, California, 1050 m depth (Lamshead and Boucher, 2003)	116	1.7 / 72	<0.1	87	94	
Porcupine Abyssal Plain, North Atlantic, 4850 m depth (Lamshead and Boucher, 2003)	119	1.8 / 54	<0.1	43	64	
Madeira Abyssal Plain, North Atlantic, 4950 m depth (Lamshead and Boucher, 2003)	72	1.6 / 29	<0.1	0.51		

Note: Each non-significant difference between a model and the data at a *P*-value of 5% is marked in **bold**. GM = Gambin model, LN = lognormal, LS = log-series, ZSM = zero-sum model.

* Older publications provided actual species abundance data, whilst for our own more recent data sets, the publications provide descriptions of sampling regimes, and other aspects of the ecological circumstances of the study system, but not always the raw data. These data sets are available on request from the authors.

+ These data are described in two papers: O'Dea and Whittaker (2007) and O'Dea *et al.* (2006).

factors have a low inherent dimensionality of niche space compared with systems responsive to multiple bundles of factors, and this variation in dimensionality is what is being captured in the α parameter. This is exemplified by the Ecuadorian montane bird data, where more recently disturbed forest and edge habitats have higher α -values than pristine forest and the lowest values are for open agricultural habitats [Table 1; and see community descriptions in O'Dea and Whittaker (2007)].

Hence, in essence we regard Gambin as a niche exploitation model that quantifies how resources are divided among co-existing species. By capturing the shape of the species abundance curve in a single parameter, α provides a measurement of the complexity or dimensionality of the 'bundles' of processes structuring a community. Or, alternatively, α can be thought of as a measure of the relative ability of the species making up the sample to 'reach' into the axes of the niche-space hypervolume characterizing the sampling space. Thus communities that are driven by many processes, or that are responsive to multiple gradients, will have a high α (relatively few rare species leading to a lognormal-shaped SAD), and those driven by a few overriding processes will have a low α (relatively many rare species leading to a log-series shaped curve). This ecologically based explanation for the tendency towards lognormality is similar to that given by Preston (1962), but our use of a gamma distribution extends Preston's ideas into a broader model. Similarly, some niche hierarchy models have adopted a conceptually similar approach (Sugihara, 1980; Tokeshi, 1993; Sugihara *et al.*, 2003). However, niche hierarchy models are not as flexible as Gambin, since decision rules have to be made to determine how to split niche space.

Applying these principles to the interpretation of the results presented herein (Table 1), we suggest that the higher dimensionalities found in terrestrial assemblages could indicate that they are structured by more complex suites of processes than marine assemblages. Only one marine data set, Irish Sea nematodes, reaches an α of 4.0. Within the marine data, the macrofaunal benthic data and the meiofaunal Nematoda show a similar range of values of α . By extension, we might predict that polluted assemblages will show a lower α than unpolluted communities, as the assemblage is being 'forced' by a single structuring agent, the pollutant. This is what was found in Pearson's (1975) Loch Linnhe data, where the contaminated macrofauna assemblage showing a lower α (2.9) than the control data (3.7). Also consistent with this hypothesis is the reanalysis of Lamshead and Boucher's (2003) data, which show coastal nematode data to have a higher dimensionality ($\alpha = 3.2\text{--}4.0$) than their deep-sea counterparts ($\alpha = 1.6\text{--}1.8$).

The Gambin and the zero-sum multinomial (ZSM) performed similarly well: a satisfying result when considering that a three-variable model would normally be expected to outperform a functionally equivalent one-variable model. We also noted that the ZSM uses the lack of a constraint on S (the number of species) to calculate a distribution that gets the shape flatter and a better fit by reducing S compared with the data, especially on large data sets. The ZSM model improves fit by tilting the distribution down on the left tail and up on the right tail without incurring a fit penalty as it has already truncated the right tail. This, in addition to the three variables, makes it a flexible model. Despite this flexibility, we believe that Gambin is a preferable solution on both theoretical grounds (it has fewer variables and does not alter S to achieve a fit) and practical grounds (it is computationally simpler).

We conclude that the Gambin model outperforms the lognormal, log-series, and zero-sum multinomial models as a descriptor of species abundance distributions. On grounds of parsimony, we suggest that it is more satisfactory than currently favoured models of the SAD, and that its biological meaning therefore warrants further investigation. We suggest

that the model provides a potentially useful tool for investigating biological information about the dimensionality, or complexity, of the processes structuring assemblages. In that respect, α could be considered as a type of equitability-based diversity measure with potential applications for the assessment of ecosystem 'health'.

REFERENCES

- Butler, B.J. and Chazdon, R.L. 1998. Species richness, spatial variation and abundance of the soil seed bank of a secondary tropical rainforest. *Biotropica*, **30**: 214–222.
- Chave, J. 2004. Neutral theory and community ecology *Ecol. Lett.*, **7**: 241–253.
- Currie, D.R. and Parry, G.D. 1999. Changes to benthic communities over 20 years in Port Philip Bay, Victoria, Australia. *Mar. Pollut. Bull.*, **38**: 36–43.
- Diserud, O.H. and Engen, S. 2000. A general and dynamic species abundance model, embracing the lognormal and the gamma models. *Am. Nat.*, **155**: 497–511.
- Eggeling, W.J. 1947. Observations on the ecology of the Budongo Rain Forest, Uganda. *J. Ecol.*, **34**: 20–87.
- 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. *J. Anim. Ecol.*, **12**: 42–58.
- Gaston, K.J. and Blackburn, T.M. 2000. *Pattern and Process in Macroecology*. Oxford: Blackwell Science.
- Gray, J.S., Clarke, K.R., Warwick, R.M. and Hobbs, G. 1990. Detection of initial effects of pollution on marine benthos – an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar. Ecol. Prog. Ser.*, **66**: 285–299.
- Gray, J.S., Aschan, M., Carr, M.R., Clarke, K.R., Green, R.H., Pearson, T.H. *et al.* 1998. Analysis of community attributes of the benthic macrofauna of Frierfjord-Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.*, **46**: 151–165.
- Harrod, J. 1964. The distribution of invertebrates on submerged aquatic plants in a chalk stream. *J. Anim. Ecol.*, **33**: 335–348.
- Hubbell, S.P.A. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.
- Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.*, **22**: 415–427.
- Lambhead, P.J.D. 1986. Sub-catastrophic sewage and industrial waste contamination as revealed by marine nematode faunal analysis. *Mar. Ecol. Prog. Ser.*, **29**: 247–260.
- Lambhead, P.J.D. and Boucher G. 2003. Marine nematode deep-sea biodiversity – hyperdiverse or hype? *J. Biogeogr.*, **30**: 475–485.
- MacArthur, R.A. 1972. *Geographical Ecology*. New York: Harper & Row.
- Magurran, A.E. 2004. *Measuring Biological Diversity*. Oxford: Blackwell.
- Magurran, A.E. 2005. Species abundance distributions: pattern or process? *Funct. Ecol.*, **19**: 177–181.
- Magurran, A.E. and Henderson, P.A. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**: 714–716.
- May, R.M. 1975. Patterns of species abundance and diversity. In *Ecology and Evolution of Communities* (M.L. Cody and J.M. Diamond, eds.), pp. 81–120. Cambridge, MA: Belknap Press.
- McGill, B.J. 2003. A test of the unified neutral theory of biodiversity. *Nature*, **422**: 881–885.
- McGill, B.J., Maurer, B. and Weiser, M. 2006. Empirical evaluation of neutral theory. *Ecology*, **87**: 1411–1423.
- Motomura, I. 1932. On the statistical treatment of communities. *Zool. Mag. Tokyo*, **44**: 379–383.
- O'Dea, N. and Whittaker, R.J. 2007. How resilient are Andean montane forest bird communities to habitat degradation? *Biodivers. Conserv.* (DOI: 10.1007/s10531-006-9095-9).
- O'Dea, N., Whittaker, R.J. and Ugland, K.I. 2006. Using spatial heterogeneity to extrapolate species richness: a new method tested on Ecuadorian cloud forest birds. *J. Appl. Ecol.*, **43**: 189–198.

- Pearson, T.H. 1975. The benthic ecology of Loch Linnhe and Loch Eil, a sea loch system on the west coast of Scotland. IV. Changes in the benthic fauna attributable to organic enrichment. *J. Exp. Mar. Biol. Ecol.*, **20**: 1–41.
- 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. *Ecology*, **43**: 185–215, 410–432.
- Shackell, N.L. and Franck, K.T. 2003. Marine fish diversity on the Scotian Shelf, Canada. *Aquat. Conserv.*, **13**: 305–321.
- Southwood, T.R.E. 1996. The Croonian lecture 1995 – Natural communities, structure and dynamics. *Phil. Trans. R. Soc. Lond. B*, **352**: 1113–1129.
- Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.*, **116**: 770–787.
- Sugihara, G., Bersier, L.F., Southwood, T.R.E., Pimm, S.L. and May, R.M. 2003. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl. Acad. Sci. USA*, **100**: 5246–5251.
- Tokeshi, M. 1993. Species abundance patterns and community structure. *Adv. Ecol. Res.*, **24**: 111–186.
- Volkov, I., Banavar, J.R., Hubbell, S.P. and Maritan, A. 2003. Neutral theory and relative species abundance in ecology. *Nature*, **424**: 1035–1038.
- Whittaker, R.H. 1975. *Communities and Ecosystems*, 2nd edn. New York: Macmillan.
- Williams, C.B. 1964. *Patterns in the Balance of Nature and Related Problems of Quantitative Biology*. London: Academic Press.