

Species abundance patterns: the problem of testing stochastic models

LOUIS-FÉLIX BERSIER* and GEORGE SUGIHARA

Scripps Institution of Oceanography 0202, University of California at San Diego, La Jolla, CA 92093, USA

Summary

1. Tokeshi (1990) proposed a goodness-of-fit test to distinguish among a general class of stochastic species abundance models using field data. This test is a good first step, but it is impaired by some shortcomings: there is no adjustment for the number of species (the rejection rate increases with the number of species in a data set); the variance of the generated abundance distributions are not taken into account (data sets with higher variance than a model are not rejected).

2. We propose an alternative Monte Carlo test that overcomes these problems. This is a versatile test that can be used with any stochastic model generating distributions.

Key-words: goodness-of-fit test, Monte Carlo methods, community structure, species diversity.

Introduction

In two recent papers, Tokeshi (1990, 1993) discusses an interesting class of models for generating abundance distributions that are based on the theme of sequential breakage (Poole 1974; Pielou 1975; Sugihara 1980). He also proposed a goodness-of-fit test for distinguishing among the various alternative models. Such a test would allow one to determine which, if any, of the models best describes the observed abundances from a community. Although this is an important contribution, there are some shortcomings in this test which can produce misleading results. Herein, we propose an alternative method for testing such models that overcomes these problems, and illustrate the method with an example.

Sequential apportionment models

The specific models that Tokeshi discusses are typical resource apportionment models in that they are based on the assumption that the fraction of the niche space apportioned to a species is proportional to its abundance (see MacArthur 1957; Pielou 1975; May 1975, 1981; Southwood 1978; Sugihara 1980, 1982; Magurran 1988). The niche space is sequentially divided, each new species entering the community taking a portion of this space. An apportionment rule deter-

mines the size of this portion. In Tokeshi's models, the size is not fixed by a constant or a free parameter, but is determined by a random number with a uniform distribution. This feature is aimed to accommodate sampling as well as natural variation (Tokeshi 1990, 1993). The expectation of such a stochastic model is the mean of many simulations, and a frequency distribution can be generated for each rank (Fig. 1).

The models differ in the rules used to define which fraction of the niche is split (the breakage rules). The smallest fraction is always chosen to be split further in the dominance-preemption (DP) model. The largest fraction is always chosen to be split further in the dominance-decay (DD) model. The fraction to be split is chosen at random (independently of the size of the fraction) in the random-fraction (RF) model. The probability that a fraction is chosen to be split is proportional to its size in the MacArthur-fraction (MF) model. The MF model is a sequential version of the broken-stick model (MacArthur 1957). In the random-assortment (RA) model, each fraction is fixed by a uniform random number. In the composite model (CM), the community is divided into the abundant species, which follow an assembly rule according to the DP, RF, MF or DD model, and the less common species, which represent a random assortment.

Model testing

The usual way of testing the goodness-of-fit of an observed to a theoretical distribution is to look at

*Present address: Institute of Zoology, Rue Emile-Argand 11, CH-2007 Neuchâtel, Switzerland.

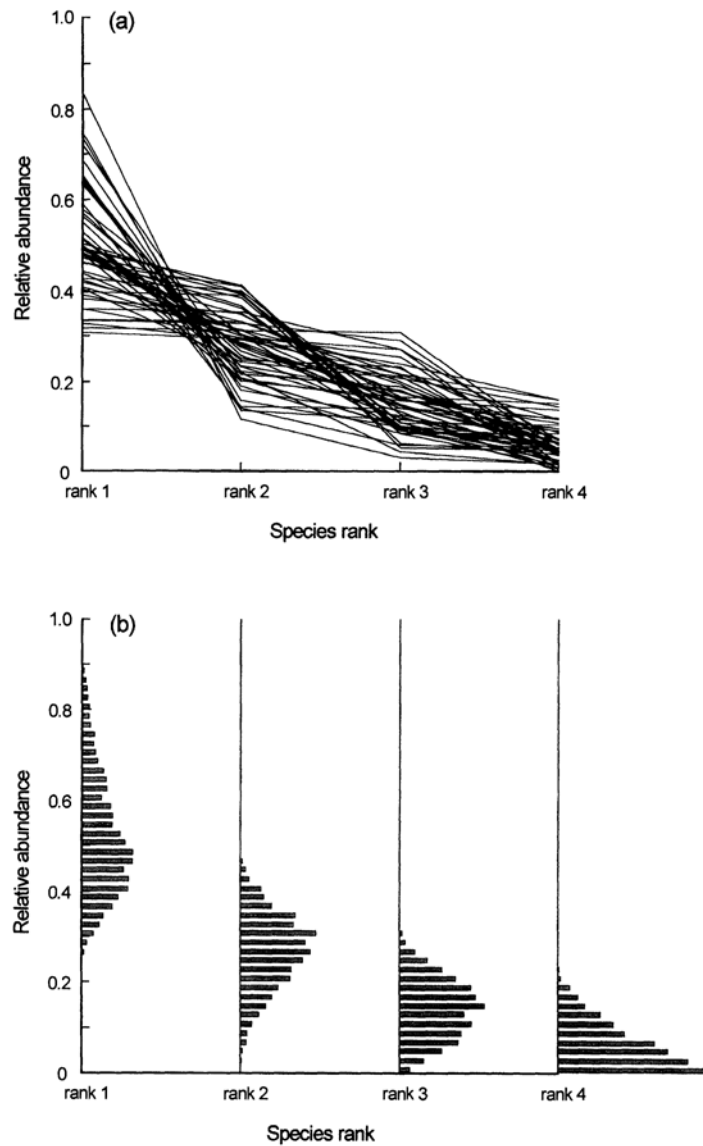


Fig. 1. An example showing the generation of distributions from a stochastic model. (a) 100 simulations of the MacArthur fraction model. (b) A frequency distribution can be constructed for each rank (the distributions shown are built from 2000 simulations).

the distributions as a whole, for example using the Kolmogorov–Smirnov test. In our specific case, the models do not produce one fixed distribution, but a collection of distributions. It is of little interest to fit a single observed abundance distribution from a community to such models, since the discriminating power between models would be very poor (Eberhard 1969; Pielou 1975; Sugihara 1980; Tokeshi 1990; Wilson 1991). Instead, one needs replicated abundance distributions from that community, treating each field distribution as equivalent to a separate model realization. Replicated tests of goodness-of-fit (Sokal & Rohlf 1981) allow one to compare replicated observed distributions with a single theoretical one. However, no standard statistical test is available for a collection of theoretical distributions that are generated by stochastic models like those considered here.

The analytical technique Tokeshi (1990) adopts is to compare the mean of each rank (m_i , with $i = 1$

to S , the number of species in a community) of the theoretical distribution and of replicated observations. For a particular model, he computes a ‘parent population’ of 10 000 simulations. Then, for each rank, he computes 95% confidence limits, given that a random sample of size n is drawn from the parent distribution (n corresponds to the number of observed replicated distributions). He accepts the model if the mean of the observed abundances of each rank falls within the corresponding confidence limits.

This test is a good first step, however, there are some problems with this method.

1. Although one major feature of these models is their stochasticity, the test disregards the heterogeneity of both the data and the model. If observed replicated distributions from a community have a much higher scatter than a particular model, but the same mean, mechanisms other than those implied by the models

may be operating (e.g. the replicates may come from communities with different forms of organization). Tokeshi's test does not allow the rejection of datasets whose variance is larger than what would be expected from a model.

2. The test is sensitive to the number of species in the community. Because all the mean observed abundances have to fall in the corresponding confidence limits in order for a model to be accepted, a distribution with many ranks is more prone to rejection than a distribution with few ranks.

3. Some distributions of ranks are skewed (especially those of the lowest ranks, see Figs 1 and 2); this renders the use of symmetric confidence limits inappropriate.

4. A last problem comes from the lack of independence among ranks, which is not taken into account in the test.

A way to remedy these problems is to make a test based on the Monte Carlo method (Manly 1990; Good 1993). With n observed replicated distributions, one could generate distributions of the means and of the variances for each rank by computing, for the model of interest, $x \times n$ simulations (x typically being 999 or more). For each set of n simulations, the mean M_i and variance V_i of each rank are computed and frequency distributions of these values are constructed. One could then test the goodness-of-fit by inspecting where the observed mean and variance, for each rank, fall in the corresponding distributions. This procedure resolves the problem of the skewed distributions (point 3 above); inspecting the variances settles the problem of heterogeneity of the observed replicates (point 1 above). However, a persisting problem here concerns the fact that one should examine each rank separately (as in Tokeshi's test), which renders the test sensitive to the total number of species in the community (point 2 above). The use of a global statistic for all the ranks will overcome this limitation, as well as the problem of the lack of independence among the ranks (point 4 above).

A solution is the following: from the procedure above, the probability that an observed mean m_i (or variance v_i) is predicted by a particular model can be estimated by inspecting its position in the distribution of M_i (or of V_i). This estimated probability p for the rank i is

$$p_i = \min \left[\frac{2s_i}{x+1}, \frac{2l_i}{x+1} \right], \quad \text{eqn 1}$$

where s_i is 1 plus the number of theoretical means M_i (or variances V_i) smaller than the observed value m_i (or v_i), and l_i is 1 plus the number of theoretical means M_i (or variances V_i) larger than the observed value m_i (or v_i) (s_i and l_i are multiplied by 2 since it is a two-tailed test; 1 is added to x since the observed value is included in the theoretical distribution). We can use

Fisher's (1954) formula as a global statistic to combine the p_i of all ranks:

$$T = -2 \sum \ln p_i \quad \text{eqn 2}$$

The observed T value (T_{obs}) is a measure of how well the model fits the data (the smaller the value the better the fit). T_{obs} can then be compared with expected T values in the following manner: (i) choose randomly n simulations from the model and compute p_i and T ; (ii) repeat this procedure many times ($y = 999$ or more); (iii) find the position of T_{obs} in the distribution of theoretical T values. The final p value is given as

$$p = \frac{l}{y+1} \quad \text{eqn 3}$$

where l is 1 plus the number of theoretical T values larger than T_{obs} (one-tailed test). This p is an estimate of the probability that T_{obs} is drawn from T values generated by the model or, in other words, the probability that the means m_i (or the variances v_i) of the observed distributions fit the model.

The whole procedure is thus composed of two main steps: (i) computation of the observed T_{obs} , and (ii) computation of the distribution of theoretical T values. This is done for the means and the variances. If the means of the observed distributions fit the model of interest (at a predefined significance level), the variances need to be tested to check if the observations show a higher (or lower) heterogeneity than that predicted by the model. Higher moment statistics of the distributions (skewness, Kurtosis) can be tested by the same procedure. For simplicity, however, these additional tests have not been implemented here.

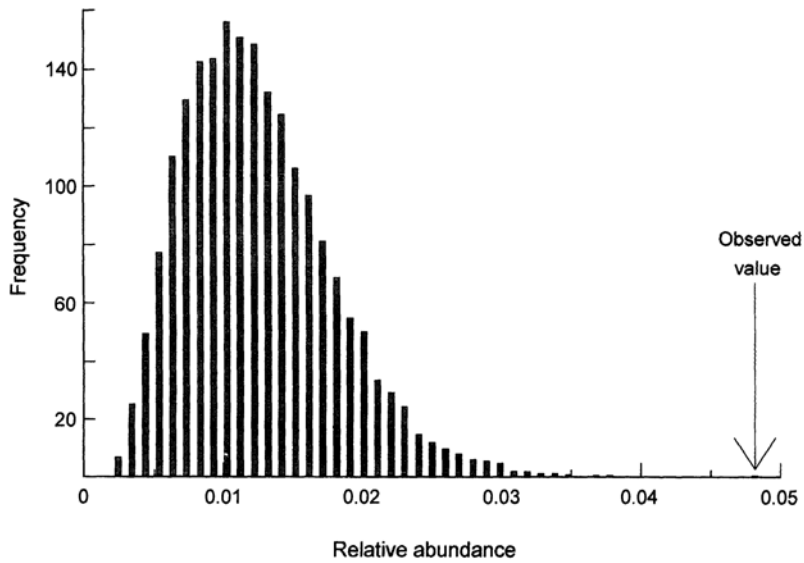
Example

To illustrate the test procedure, data from a bird census in riparian forests in Brittany are used here (Bersier & Meyer 1994). The abundances of four birds, the chiffchaff *Phylloscopus collybita*, the blackcap *Sylvia atricapilla*, the garden warbler *Sylvia borin* and the long-tailed tit *Aegithalos caudatus* were estimated in a total of 15 plots of 100 m wide and about 500 m long at four different locations. We have randomly chosen one plot from one location and two non-adjacent plots from the other three locations. The relative abundances of the chosen species in these seven plots are shown in Table 1.

For each model, we have performed 1999 times 7 simulations, and computed the means and variances for each rank of the 1999 theoretical samples. The distribution of the means of the fourth rank for the DP model is shown in Fig. 2. The observed relative abundance ($m_4 = 0.048$) is the largest value of the distribution, which leads to an estimated p_4 of 0.001. Table 2 presents the p_i values for the means and variances of each rank for the six models. From these p_i values, the T_{obs} are computed for each model using

Table 1. Relative abundance of four bird species. The sample is composed of seven plots located in four different forests (A to D)

Plot	<i>Phylloscopus collybita</i>	<i>Sylvia atricapilla</i>	<i>Sylvia borin</i>	<i>Aegithalos caudatus</i>
A2	0.591	0.341	0.029	0.039
B1	0.440	0.275	0.175	0.109
B3	0.624	0.294	0.021	0.062
C1	0.520	0.288	0.148	0.044
C3	0.528	0.443	0.023	0.007
D2	0.596	0.308	0.050	0.047
D4	0.568	0.344	0.059	0.047
Rank	1	2	3	4
Mean	0.552	0.327	0.072	0.048
Variance	0.00383	0.00327	0.00402	0.00102

**Fig. 2.** Frequency distribution of the mean abundance of the fourth rank from the DP model ($x = 2000$ simulations; $n = 7$ replicated observations). The observed mean ($m_4 = 0.048$) is the largest value of the distribution, resulting in $p_4 = 0.001$ (Table 2).**Table 2.** Estimated p_i values for the observed means (\bar{M}) and variances (s^2) of each rank i of the distributions with the six models of niche apportionment. The first two species of the composite model follow the RF model, the last two the RA model. T_{obs} is computed from these values using Eq. 2. The final p value is determined by comparing the T_{obs} with T values obtained from simulations of the model itself

	Random fraction		Random assortment		Composite model		MacArthur fraction		Dominance decay		Dominance preemption	
	\bar{M}	s^2	\bar{M}	s^2	\bar{M}	s^2	\bar{M}	s^2	\bar{M}	s^2	\bar{M}	s^2
p_1	0.400	0.013	0.648	0.018	0.171	0.002	0.528	0.079	0.044	0.240	0.001	0.010
p_2	0.025	0.077	0.031	0.094	0.004	0.057	0.061	0.349	0.178	0.716	0.006	0.041
p_3	0.203	0.640	0.136	0.543	0.368	0.697	0.002	0.926	0.001	0.950	0.131	0.155
p_4	0.834	0.536	0.705	0.342	0.941	0.500	0.414	0.340	0.150	0.153	0.001	0.005
T_{obs}	12.74	15.94	12.53	16.19	16.51	20.48	21.27	9.50	27.53	7.37	42.32	30.48
p	0.1725	0.0665	0.1595	0.0535	0.0805	0.0240	0.0205	0.2910	0.0055	0.4575	0.0005	0.0010

equation 2. Then the distributions of theoretical T s are built. We computed 1999 T s for the means and the variances of each model (this involved repeating 1999 times 1999 simulations). The distributions of T s for

the means and the variances of the composite model are shown in Fig. 3. For the means, 160 T s from the model are larger than T_{obs} , which leads to a global p of $161/2000 = 0.0805$ (Fig. 3a, Table 2). This indicates

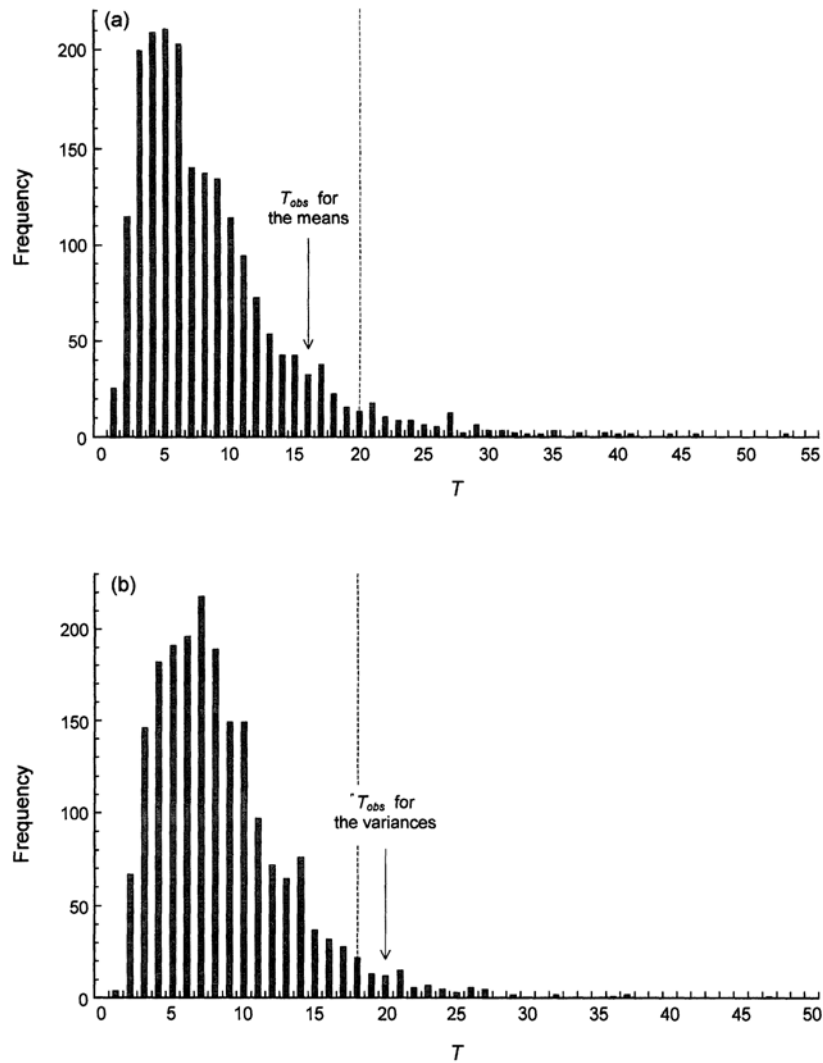


Fig. 3. Frequency distributions of the T values for (a) the means, and (b) the variances of the composite model (2000 simulations). The dashed line indicates the limit of the 5% significance level. The means of the 7 observed distributions fit the model, but their variance is different from that predicted by the model.

that, using a significance level of 0.05, the means of the seven observed distributions fit the CM model. However, these observed distributions can show a greater scatter than that predicted by the model, which is tested by inspecting the variances. Only 47 T values from the model are larger than T_{obs} for the variances, which produces a global p of 0.024 (Fig. 3b, Table 2). Thus, one must reject the hypothesis that the heterogeneity of the observed distributions is the same as that predicted by the CM model. From Table 2, it is apparent that the observed distributions fit the RA, and RF models, the last one giving a slightly better fit.

Comments

The proposed test is computer intensive. In this example, computing the distribution of T s for one model involved about 4 000 000 permutations (it took about 2 hours on a pentium-based computer). However, once such a distribution computed, it can be

used to test other data sets, as long as the number of species and of replicated observed distributions remain the same.

This test is versatile in that it could be used with any stochastic model generating distributions. In the case of species-abundance models, it could be used to test field data with the power fraction model (Tokeshi 1996) and with the niche hierarchy model (Sugihara 1980). The power fraction model possesses a free parameter k that allows a tuning of the breakage rule. The value T_{obs} for the means could be used as a criterion to find k (the best k is the one that minimizes T_{obs}). Then, the global significance test proposed here could be carried out by generating the distribution of T values from the model with the particular parameter k .

As originally stated, the computation of the niche hierarchy model requires no stochastic component in the apportionment rule: the fragment chosen to be split is broken in two portions of size 0.25 and 0.75 of the original fragment (Sugihara 1980). However, this fixed ratio was used as an approximation of a tri-

angular distribution found empirically in two-species assemblages (Fig. 4 in Sugihara 1980). As such, it could be envisaged to use random numbers drawn from a triangular distribution to generate the niche hierarchy model, and to use the test proposed here with field data. It must be noted that a triangular distribution with a peak at 0.75 will not produce the canonical lognormal distribution (Sugihara 1982; Siegel & Sugihara 1983), but that a peak at 0.66 is compatible with the canonical hypothesis.

In the case of the specific models considered here, we suggest that it is necessary to check the heterogeneity aspect by examining the variances. This is because the stochasticity is an important feature of Tokeshi's models. However, there possibly exist stochastic models where heterogeneity need not be tested. This is the decision of the investigator. In contrast, there could be an interest not only in the dispersion, but also in the asymmetry and in the peakedness/flatness of the distributions of each rank. These additional controls of the skewness and Kurtosis are easily implemented with the proposed test.

In summary, the Monte Carlo test proposed here can be easily implemented to test the fit of stochastic models of species abundance with empirical data. This test uses the actual distributions of a model, and takes into account the heterogeneity of the observations and of the model. We believe it represents a substantial improvement over previous tests and will be useful in future studies of species abundance patterns.

Acknowledgments

We are grateful to Ian Abramson, Daniel Borcard, Stuart Hurlbert, Cleridy Lennert, Dietrich Meyer, Maria Milicich, Peter Perkins, Richard Penner and Mutsunori Tokeshi for their help. We are indebted to an anonymous referee who suggested the use of the Fisher formula as a global statistic. This work was supported by the Swiss National Science Foundation, by Office of Naval Research Grants N00014-92-J-4068 and N00014-95-1-0034, by endowment income from the John Dove Isaacs Chair in Natural Philosophy, and by the Ciba-Geigy Jubilee Foundation.

References

- Bersier, L.-F. & Meyer, D.R. (1994) Bird assemblages in mosaic forests: the relative importance of vegetation structure and floristic composition along the successional gradient. *Acta Oecologica*, **15**, 561–576.
- Eberhardt, L.L. (1969) Some aspects of species diversity models. *Ecology*, **50**, 503–505.
- Fisher, R.A. (1954) *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.
- Good, P. (1993) *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer-Verlag, New York.
- MacArthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Sciences of the USA*, **43**, 293–295.
- Magurran, A.E. (1988) *Ecological Diversity and Its Measurement*. Princeton University Press, Princeton, NJ.
- Manly, B.F.J. (1990) *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall, New York.
- May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and Evolution of Communities* (eds M. L. Cody & J. M. Diamond), pp. 81–120. Belknap/Harvard University Press, Cambridge, MA.
- May, R.M. (1981) Patterns in multi-species communities. *Theoretical Ecology: Principles and Applications* (ed. R. M. May), pp. 197–227. Blackwell Scientific, Oxford.
- Pielou, E.C. (1975) *Ecological Diversity*. Wiley-Interscience, New York.
- Poole, R.W. (1974) *An Introduction to Quantitative Ecology*. McGraw-Hill, New York.
- Siegel, A.F. & Sugihara, G. (1983) Moments of particle size distributions under sequential breakage with applications to species abundance. *Journal of Applied Probability*, **20**, 158–164.
- Sokal, R.R. & Rohlf, F.J. (1981) *Biometry*. Freeman and Co., San Francisco.
- Southwood, T.R.E. (1978) *Ecological Methods*. Chapman & Hall, London.
- Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *American Naturalist*, **116**, 770–787.
- Sugihara, G. (1982) *Niche hierarchy: structure, organization, and assembly in natural communities*. PhD Dissertation, Princeton University, Princeton, NJ.
- Tokeshi, M. (1990) Niche apportionment or random assortment: species abundance patterns revisited. *Journal of Animal Ecology*, **59**, 1129–1146.
- Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, **24**, 111–186.
- Tokeshi, M. (1996) Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos*, **75**, 543–550.
- Wilson, J.B. (1991) Methods for fitting dominance/diversity curves. *Journal of Vegetation Science*, **2**, 35–46.