ORIGINAL ARTICLE

Some species flourish when many do not: a pattern in data on ecological communities

Gerben Straatsma[1] · Edwin T. H. M. Peeters[1] · Egbert H. van Nes[1]

Abstract

Patterns in species × sample tables of communities depend above all on the organisms of the data sets and the conditions involved. Patterns that surpass individual sets are of special interest. Our question, looking for a shared pattern in 12 sets, is if relative abundances among species are independent of the sample, or formulated alternatively, if species have abundances that are correlated with total abundances over samples. For exploration we study the overdispersion/aggregation of the data. A relatively high variation in the total abundances of samples is noticed, indicating an effect of environmental variation. Overdispersion imposes constraints on the accommodation of relatively high abundance values to samples with a relatively low total abundance. The null hypothesis of 'no association' is modelled by permutation/resampling of the data at the level of the individual. A correlation study of actual and permuted sets is performed. All actual sets contain a significant number of species that defy our question. These species flourish when many do not. The relation of our question with issues in theoretical ecology, such as the assumption of a neutral effect of environmental conditions and/or of neutral characteristics of species, is discussed.

Keywords Overdispersion · Permutation · Macroecology · Neutrality · Assembly theory

Introduction

Patterns in multi-sample data sets of communities depend on the organisms of the data sets and the conditions involved. Taxonomical characteristics and environmental conditions separate the sets. We are interested in patterns that are shared by different sets, as Lawton (1999) and McGill (2019), ignoring specific properties of species and samples. The existence of such a pattern, a law, is relevant for theories and models for ecological community assembly. These theories (Chase & Myers, 2011) are falsified if they contradict a law.

Our initial interest was raised by a long-term data set on mushrooms (Egli et al., 1977; Straatsma et al., 2001). Mushrooms are sporocarps of, mostly, long living mycelia in soil (Cairney, 2005). Since the dynamics in the soil are low annually, we wondered if years with relatively few individuals were dominated by the same species as years with

relatively many. This question can be articulated as: are relative abundances among species independent of the sample, or formulated alternatively, do species have abundances that are correlated with total abundances over samples? This can be asked for multi-sample data sets of any community. We could not find any (macro)ecological literature addressing this question. We study 12 data sets, as species × sample tables (also called cross or contingency tables), that cover different biological groups, in different regions on earth, in different ecosystems.

Ecological data are generally 'overdispersed' (Bliss & Fisher, 1953). The term is used if the variance of values exceeds their mean. It captures the aggregated nature of individuals, implying a relatively high chance of cells in the tables to hold no individuals at all (Warton, 2005). We anticipate that overdispersion is relevant for our analysis and study it as a matter of statistical exploration (Zuur et al. 2010). From population studies it is well known that the variance of abundance increases with the mean, showing a straight line in a log/log plot (Taylor's law on 'fluctuation scaling'; Taylor, 1961; Brown et al., 2017). We wonder if variance and mean within species and within samples also produce this pattern. We also wonder what overdispersion/

✉ Gerben Straatsma
   gerben.straatsma@gmail.com

1   Department of Environmental Sciences, Wageningen
    University and Research, Droevendaalsesteeg 3,
    6708 PB Wageningen, The Netherlands

aggregation means for the sample with the highest total abundance and for the highest abundance values of species. Is the 'capacity' of this sample sufficient to hold all highest abundances of all species, as in a table with expected values.

The analysis of structure in contingency tables is a long- and well-known topic in statistics (Stigler, 2002). For a null model we can use the ordinary procedure and produce a table with expected values, calculated from the marginal totals (the total abundances of species and of samples), to be tested against the actual values with a chi-squared test. Alternatively, data can be resampled/permutated at the level of the individual (Capone & Kushlan, 1991; see also Ulrich & Gotelli, 2010). This is the method of our choice, as it acknowledges the discrete nature of the data.

Actual and modelled tables are analysed for overdispersion and for correlations between the abundance of species and total abundance, over samples. Additionally, the number of species showing their highest abundance in the sample with the highest total abundance is determined.

Material and methods

Data sets

Twelve multi-sample data sets, each with a constant sampling effort, are analysed. We acknowledge the biocurators in their efforts to collect the data, in publishing and/or sharing them, in Table 1. Data sets cover different biological groups, different regions on earth in different ecosystems. The number of observed individuals range from 14,965 to 1,178,380. The number of species is highly variable among the datasets with a minimum of 16 species in set 3 and a maximum of 502 in set 11. We used the sets for other pattern

analyses already, on rarity (Straatsma & Egli, 2012) and on the species abundance distribution (Straatsma & Egli, 2017).

Null hypothesis and model

The choice of our null model starts with a thought exercise. The simplest null model assigns all individuals randomly to the cells of the species × sample table. We can do that using an auxiliary table with the number of rows equal to the number of individuals and three columns, the first for the individual's number, the others for tags of species identity and of sample code, respectively. Tags for species and samples can be randomly assigned and a species × sample table constructed. Cell values will approximate a Poisson distribution, characterized by values whose variation equals the mean. The same is true for the totals of species and of samples (the marginal totals); they can either be totalled over cell values or determined directly from the individual's tags for species or for samples, respectively. Poisson distributions contradict ecological facts. Ecological data are generally 'overdispersed' (Bliss & Fisher, 1953). (Total) abundances of species follow a law that is known for over a century (McGill et al. 2007; 'species abundance distribution'): 'every community shows a hollow curve or hyperbolic shape on a histogram with many rare species and just a few common species'. We are not aware of a law or pattern for total abundances of samples.

For an appropriate null model, we should acknowledge the total abundances of the species as they are. Further, we need to know if there is more to the total abundances of samples than randomness. In the results section, we will learn that their variation exceeds the mean of sample totals, in all data sets. Thus, we also should acknowledge these

Table 1 Set origins and characteristics. The availability of the data is mentioned in a note at the end of the text. Abundances are number of individuals (organisms in most sets, colonies in set 9)

Data set		Origin of data	Total abundance	Number of samples	Number of species
1	Mushrooms	Egli et al. 1977, 2006	108,014	551	408
2	Fish	Henderson & Bird, 2010	143,420	357	83
3	Crustaceans	Ibid	1,039,322	340	16
4	Fish + Crustaceans	Ibid	1,178,380	340	99
5	Trees	Condit et al. 1996, Hubbell et al. 2005	206,513	512	295
6	Rodents	Brown et al. 2001, Ernest et al. 2009	33,013	617	40
7	Winter annuals	Ibid	414,080	282	47
8	Summer annuals	Ibid	364,958	335	49
9	Ant colonies	Ibid	14,965	288	20
10	Seedlings	Firbank et al. 2003	284,678	915	240
11	Brachiopod fossils	Cooper & Grant, 1977, Olszewski & Erwin, 2004	793,077	187	502
12	Flies	Papp, 2007	92,680	1,015	106

totals as they are. We arrive at the analysis of structure in contingency tables with given marginal totals, a long- and well-known topic in statistics (Stigler, 2002).

For contingency tables with given marginal totals, expected values of the table cells can be calculated by the product of the total abundance of the species involved and of the total abundance of the sample involved, divided by the total abundance of the set. In a table with expected values, all species of a data set will show a perfect correlation between their abundance values and the total abundances of the samples. Also, all species of a data set will show their highest abundance in the sample with the highest total abundance.

The calculation of expected values results in decimal values, not discrete ones as the counts in the actual sets. This problem can be met with resampling/permutation of the data at the level of the individual. Again, we can use an auxiliary table, now for every individual of an actual species×sample table, with two tags already for species and sample. These tags can be re-shuffled, and a permuted table obtained (Capone & Kushlan, 1991; see also Ulrich & Gotelli, 2010, their 'IT' algorithm). This represents our null model. The results of permutation will vary due to chance placement of individuals in table cells. Nevertheless, cell values will approximate expected values. To account for chance placement, we use 100 permutation runs in our final analysis.

Overdispersion/aggregation

Overdispersed distributions have also been called 'contagious', indicating that the presence of one individual in a cell of a table increases the chance of there being some more (terminology of G. Pólya; in Bliss & Fisher, 1953). Aggregation of individuals implies a relatively high chance of cells to hold no individuals at all, that the distribution is 'zero-inflated' [Warton, 2005; a distribution is truly zero-inflated only if not even the negative binomial distribution (Bliss & Fisher, 1953) can account for the large number of zeros]. For the analysis of overdispersion, mean and variance are determined of cell values, zeros included, and of marginal totals. Additionally, the number of cells holding one or more individuals is counted.

The number of table cells, n, holding a zero value under a Poisson distribution, can be calculated by the expression:

$$n = \exp\left(-n_{\text{tot}} / \left(n_{\text{samp}} * n_{\text{spec}}\right)\right) * n_{\text{samp}} * n_{\text{spec}}$$

where n_{tot}, n_{samp} and n_{spec} represent the total abundance, the number of samples and the number of species of the set (the expression for a zero value gets simple by the circumstance that $0! = 1$).

Correlation analysis

Our question, if there is a correlation between the abundances of species in samples and the total abundance of samples, is tested by linear regression between log transformed values of both variables. Only abundances 'when present', of one or higher, are used. The number of species with a significant positive or negative correlation ($p < 0.05$) is used as a measure. Some of the species have too few individuals or too little variation to determine a reliable correlation. These species are considered not eligible for analysis.

We also determine the number of species showing their highest abundance in the sample with the highest total abundance.

The 5% and 95% percentiles of the 100 sets of permuted data are used as confidence limits. Statistical analysis is done in MATLAB; code is available from Online Resource 1.

Results

Exploratory statistics; overdispersion

The numbers of species and of samples of each data set is given in Table 1. The range of abundances of species and of samples is illustrated in Fig. 1 and shows that these ranges are different among sets and larger for abundances of species than of samples. Set 5 stands out by its relatively low level of variation between samples (also see Online Resource 2).

In all species×sample tables, the variance of the abundances of table cells exceeds the mean (Table 2a), indicating that individuals are overdispersed. The same is true for marginal totals of species and of samples (Table 2b, c). This result implies that the marginal totals should be accounted for in the null model.

The numbers of zero values of table cells (the complement of values larger than zero in Table 2d) are high in
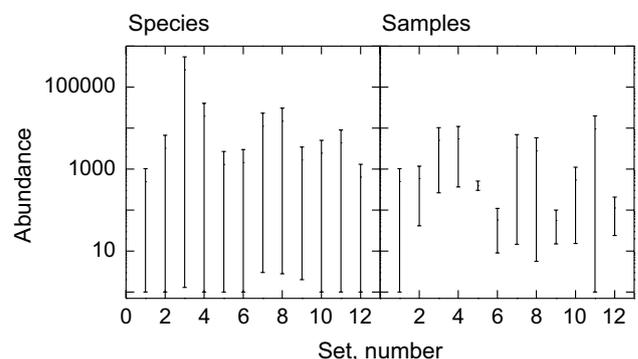

Fig. 1 Range of total abundances of species and of samples of the actual sets; ranges are 90% quantiles. Sets are numbered according to Table 1

Table 2 The extent of overdispersion; a, b, c) variance to be compared with mean, d) number of cells with an abundance value > 0 to be compared with total number of cells. Sets are numbered according to Table 1

Data set	a		b		c		d	
	Abundance, table cells		Total abundance, species		Total abundance, samples		Number of table cells	
	Mean	Variance	Mean	Variance	Mean	Variance	Total	Value > 0
1	0.480	123	265	1.08E+06	196	1.97E+05	224,808	8,865
2	4.84	3,310	1,728	5.33E+07	402	3.35E+05	29,631	5,241
3	191	1.06E+06	6.50E+04	3.04E+10	3057	1.79E+07	5,440	2,171
4	35.0	1.79E+05	1.19E+04	5.24E+09	3466	1.90E+07	33,660	7,206
5	1.37	38.1	700	6.76E+06	403	3,892	151,040	38,593
6	1.34	30.8	825	3.03E+06	53.5	1,041	24,680	4,299
7	31.2	8.48E+04	8,810	1.16E+09	1468	3.80E+06	13,254	3,902
8	22.2	8.11E+04	7,448	8.24E+08	1089	5.01E+06	16,415	3,702
9	2.60	34.1	748	1.27E+06	52.0	772	5,760	2,520
10	1.30	843	1,186	3.96E+07	311	3.13E+05	219,600	12,069
11	8.45	4.57E+04	1,580	2.20E+07	4241	1.78E+08	93,874	4,342
12	0.86	58.5	874	3.63E+07	91.3	4.38E+03	107,590	11,465
p Wilcoxon test;	0.001		< 0.001		0.002			

comparison to the number expected from a Poisson distribution (not shown; can be calculated from the expression given in Materials and Methods). We considered the negative binomial distribution (Bliss & Fisher, 1953) for total abundances. Our impression is that predicted values from the negative binomial model resemble observed values better for total abundances of samples than of species (results not shown).

The variances of abundances within species and within samples plotted against their means show the pattern known as Taylor's law (Online Resource 2).

More cells of the species × sample tables of permuted data hold individuals than the tables of the actual sets (Online Resource 3a). This implies a relaxation of overdispersion by permutation; very high values are less high, the number of species having abundance values of just one increases (Online Resource 3a).

Correlation between species abundances and total abundances of samples

The number of species with abundances significantly correlated with total abundances of samples is significantly lower in the actual sets than in the sets with permuted data (Fig. 2a; Online Resource 3b). A collateral result of permutation is that the numbers of species eligible for calculating the correlation coefficient are lower in sets of permuted data than in the actual sets; except for sets 3 and 7 (Online Resource 3b).

Sample with highest total abundance

The total abundance of the sample with the highest total abundance is lower than the sum of the highest abundances

Fig. 2 Statistics of actual and permuted sets with respect to species behaviour. Sets are numbered according to Table 1. Open dots: actual data, Bars: 90% quantile ranges for 100 permutated sets. **a**: the number of species having a significant positive relation; **b**: the number of species having their highest abundance in the sample with the highest total abundance. Data to this figure are given in Online Resource 3
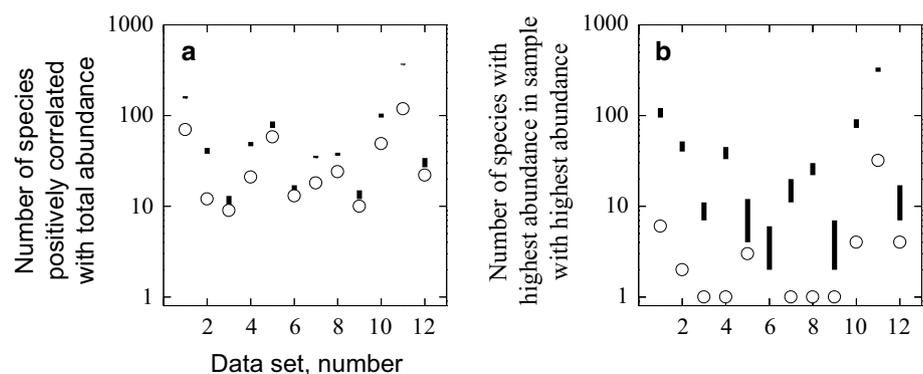
Table 3 Highest total abundance of samples and their failing capacity to accommodate the highest abundances of all species. Sets are numbered according to Table 1

Data set	Highest total abundance of samples	Sum of highest abundances of all species
1	4006	15,156
2	8060	12,718
3	38,328	53,734
4	39,312	66,420
5	629	3812
6	181	719
7	9936	21,055
8	13,943	33,241
9	159	395
10	7591	43,289
11	116,705	419,639
12	585	1195

of species (Table 3). This sample cannot accommodate all species with their highest abundances.

Significantly fewer species in the actual sets have their highest abundance in the sample with the highest total abundance than in the sets with permuted data (Fig. 2b; Online Resource 3c). Occasionally, in the actual sets, a species has its highest abundance in a sample with a relatively very low total one (not shown).

Discussion

Overdispersion

Overdispersion occurs in all sets. Although variance varies among species and samples, it followed approximately a power-law function with mean density according to Taylor's law.

Samples with a low total abundance will have difficulty to accommodate the highest abundance of a species that has a high total abundance. Possibly this single value exceeds the total abundance of such a sample. Only species with a moderate to low total abundance can have their highest abundances in samples with a moderate to low total abundance. However, not exclusively: they can also have it in samples with a high total abundance. Because not all species can have their highest abundance in the sample with the highest total abundance, some species will have that in other samples. The high abundance values rather than the low ones are constrained to samples with a high total abundance. We conclude that the high level of aggregation of individuals is one side of the coin and that, somehow, differential behaviour of species is the other.

Environmental variation among samples

Total abundances of samples in all sets are overdispersed. There is something beyond randomness that causes this high level of variation. We suggest that it is an effect of environmental variation. In our study it was logical to ignore the specific characteristics of samples and the environmental conditions involved; such characteristics separate sets, are not shared. Specific effects of environmental conditions in some of the sets have been published: seasonality in the Mushroom (Straatsma et al., 2001) and Fish (Shimadzu et al. 2013) sets; water availability in the Tree set (Engelbrecht et al., 2007) and precipitation in the sets of Rodents and Annual plants (Ernest et al., 2000). Environmental variation always plays a role, whatever the scale of sampling (Wiens, 2000). The environment is patchy; even in what is intuitively taken a homogeneous environment, water, micro-patches occur (Jenkinson et al. 2015). The effects of environmental conditions illustrate the contingency issue mentioned by Lawton (1999) and McGill (2019). The specific environmental conditions are relevant to the specific set only. What all sets share is the naked fact of variation of environmental conditions.

Correlation

Significant differences between actual and permuted sets are present in a) the number of species with a positive relation with total abundances and b) the number of species that has the highest abundance in the sample with the highest total abundance. The null hypothesis is rejected and a non-stochastic pattern in the data sets is accepted. Both, the study of overdispersion and the correlation analysis, lead to the conclusion that some species flourish when many do not. The pattern found is not surprising; it bears similarity to naturalist's observations that otherwise rare species can occur abundantly at peculiar sites like the occurrence of metallophytic plants in calaminarian grasslands [like *Viola lutea* subsp. *calaminaria* (Heimans, 1911); calamine is an ore of zinc]. The pattern has not explicitly been recognized in macroecology.

Quite some species show abundances that correlate with total abundances of samples in actual sets. That suggests a form of neutral behaviour.

Relation with theoretical ecology

One could hardly expect an affirmative answer to our question, that there is a perfect correlation between the abundances of species over samples with the total abundances of the samples. Affirmation would imply that the null hypothesis of no association would not be rejected. What would

this mean? Then, all species show an identical response, in proportion to their total abundance, to variation among samples. Does that imply that the species involved behave neutrally? Not necessarily. One can speculate that samples represent nothing else than total abundance, rejecting grip for differential behaviour of species. Only if this speculation can be rejected, neutral behaviour of species needs consideration. Thus, in case of an affirmative answer, we are confronted with two assumptions, one on neutrality of samples and one on species. This relates to a result of Clark (2012). He analysed assumptions of ecologist that either take sites or species equivalent (niche ecologists versus neutralists, respectively) and judged that models based on their seemingly opposite assumptions can yield the same answer.

As we rejected the hypothesis of no association, neither of the speculations above receives support. It implies that samples as well as species have functionally different characteristics. This conclusion will correspond with the intuition of naturalists, however, the subject of neutrality is relevant in theoretical ecology. The concept is strongly influenced by the theory of Hubbell (2001) on the replacement of individuals in communities ('births' and 'deaths'). Our question is not directly aimed at replacement. Neutrality in a general meaning relates to issues of coexistence of species and stability of communities, for which we refer to Chase and Myers (2011).

The pattern found does not stand alone. Well-known other patterns are listed in Lawton (1999) and McGill (2019). Our results on overdispersion and the constraints it poses, as well as that Taylor's law holds among species and samples, may indicate the existence of additional patterns. None of these alternative patterns should be contradicted by theories on community assembly [see Chase and Myers (2011) for theories].

The question if species have abundances that are correlated with total abundances over samples is answered negatively. Instead, we found a pattern in all data sets that some species flourish when many do not. The existence of the pattern is linked with overdispersion/aggregation of individuals. Although the pattern is perhaps not surprising, it conflicts with assumptions on neutrality of characteristics of samples and/or species, a topic in theoretical ecology.

Declarations

References

Bliss, C. I., & Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data / Note on the efficient fitting of the negative binomial. *Biometrics, 9*, 176–200. https://doi.org/10.2307/3001850

Brown, J. H., Whitham, T. G., Ernest, S. K. M., & Gehring, C. A. (2001). Complex species interactions and the dynamics of ecological systems: Long-term experiments. *Science, 293*, 643–650. https://doi.org/10.1126/science.293.5530.643

Brown, M., Cohen, J. E., & De la Pena, V. H. (2017). Taylor's law, via ratios, for some distributions with infinite mean. *Journal of Applied Probability, 54*, 657–669. https://doi.org/10.1017/jpr.2017.25

Cairney, J. W. G. (2005). Basidiomycete mycelia in forest soils: Dimensions, dynamics and roles in nutrient distribution. *Mycological Research, 109*, 7–20. https://doi.org/10.1017/S0953756204001753

Capone, T. A., & Kushlan, J. A. (1991). Fish community structure in dry-season stream pools. *Ecology, 72*, 983–992. https://doi.org/10.2307/1940598

Chase, J. M., & Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Phil Trans R Soc B, 366*, 2351–2363. https://doi.org/10.1098/rstb.2011.0063

Clark, J. S. (2012). The coherence problem with the unified neutral theory of biodiversity. *Trends in Ecology & Evolution, 27*, 198–202. https://doi.org/10.1016/j.tree.2012.02.001

Condit, R., Hubbell, S. P., Lafrankie, J. V., Sukumar, R., Manokaran, N., Foster, R. B., et al. (1996). Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology, 84*, 549–562. https://doi.org/10.2307/2261477

Cooper, G. A., & Grant, E. (1977). *Permian Brachiopods of West Texas, VI Smithsonian contributions to paleaobiology 32*. Smithsonian Institution Press.

Egli, S., Ayer, F., & Chatelain, F. (1977). Die Beschreibung der Diversitaet von Makromyzeten. Erfahrungen aus pilzoekologischen Langzeitstudien im Pilzreservat la Chaneaz. *Mycologia Helvetica, 9*, 19–32

Egli, S., Peter, M., Buser, C., Stahel, W., & Ayer, F. (2006). Mushroom picking does not impair future harvests—results of a long-term study in Switzerland. *Biological Conservation, 129*, 271–276. https://doi.org/10.1016/j.biocon.2005.10.042

Engelbrecht, B. M. J., Comita, L. S., Condit, R., Kursar, T. A., Tyree, M. T., Turner, B. L., et al. (2007). Drought sensitivity shapes species distribution patterns in tropical forests. *Nature, 447*, 80–82. https://doi.org/10.1038/nature05747

Ernest, S. K. M., Brown, J. H., & Parmenter, R. R. (2000). Rodents, plants, and precipitation: Spatial and temporal dynamics of consumers and resources. *Oikos, 88*, 470–482. https://doi.org/10.1034/j.1600-0706.2000.880302.x

Ernest, S. K. M., Valone, T. J., Brown, J. H. (2009) Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. *Ecology 90*, 1708. https://doi.org/10.1890/08-1222.1.

Firbank, L. G., Heard, M. S., Woiwod, I. P., Hawes, C., Haughton, A. J., & Champion, G. T. et al (2003) An introduction to the farm-scale evaluations of genetically modified herbicide-tolerant crops. *Journal of Applied Ecology 40*, 2–16. https://doi.org/10.1046/j.1365-2664.2003.00787.x.

Heimans, E. (1911). *Uit ons Krijtland*. Versluys.

Henderson, P. A., & Bird, D. J. (2010). Fish and macro-crustacean communities and their dynamics in the Severn Estuary. *Marine Pollution Bulletin, 61*, 100–114. https://doi.org/10.1016/j.marpolbul.2009.12.017

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Hubbell, S. P., Condit, R., Foster, R. B. (2005) Barro Colorado forest census plot data. https://repository.si.edu/handle/10088/20925)

Jenkinson, I. R., Sun, X. X., & Seuront, L. (2015). Thalassorheology, organic matter and plankton: Towards a more viscous approach in plankton ecology. *Journal of Plankton Research, 37*, 1100–1109. https://doi.org/10.1093/plankt/fbv071

Lawton, J. H. (1999). Are there general laws in ecology? *Oikos, 84*, 177–192. https://doi.org/10.2307/3546712

McGill, B. J. (2019). The what, how and why of doing macroecology. *Global Ecology and Biogeography, 28*, 6–17. https://doi.org/10.1111/geb.12855

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., et al. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters, 10*, 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

Olszewski, T. D., & Erwin, D. H. (2004). Dynamic response of Permian brachiopod communities to long-term environmental change. *Nature, 428*, 738–741. https://doi.org/10.1038/nature02464

Papp, L. (2007) A study of the cow pat Diptera on the Hortobágy, Hungary. *Folia Entomol Hung 68*, 123–135.

Shimadzu, H., Dornelas, M., Henderson, P. A., & Magurran, A. E. (2013). Diversity is maintained by seasonal variation in species abundance. *BMC Biology, 11*, 98. https://doi.org/10.1186/1741-7007-11-98

Stigler, S. (2002). The missing early history of contingency tables. *Ann Fac Sci Toulouse, 6–11*, 563–573

Straatsma, G., Ayer, F., & Egli, S. (2001). Species richness, abundance, and phenology of fungal fruit bodies over 21 years in a Swiss forest plot. *Mycological Research, 105*, 515–523. https://doi.org/10.1017/S0953756201004154

Straatsma, G., & Egli, S. (2012). Rarity in large data sets: singletons, modal values and the location of the species abundance distribution. *Basic and Applied Ecology, 13*, 380–389. https://doi.org/10.1016/j.baae.2012.03.011

Straatsma G, Egli S (2017) The stretched exponential as one of the alternatives for the power law to fit ranked species abundance data. *Matters* 14.5/30. https://doi.org/10.19185/matters.201612000009

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature, 189*, 732–735. https://doi.org/10.1038/189732a0

Ulrich, W., & Gotelli, N. J. (2010). Null model analysis of species associations using abundance data. *Ecology, 91*, 3384–3397. https://doi.org/10.1890/09-2157.1

Warton, D. I. (2005). Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics, 16*, 275–289. https://doi.org/10.1002/env.702

Wiens, J. A. (2000). Ecological heterogeneity: an ontogeny of concepts and approaches. In M. J. Hutchings, E. A. John, & A. J. A. Stewart (Eds.), *The ecological consequences of environmental heterogeneity.* (pp. 9–31). Blackwell.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x