Taking species abundance distributions beyond individuals

H. Morlon

Ethan P. White
Utah State University

R. S. Etienne

J. L. Green

A. Ostling

D. Alonso

See next page for additional authors

Follow this and additional works at: http://digitalcommons.usu.edu/biology_facpub

Recommended Citation

This Article is brought to you for free and open access by the Biology at DigitalCommons@USU. It has been accepted for inclusion in Biology Faculty Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
Authors

This article is available at DigitalCommons@USU: http://digitalcommons.usu.edu/biology_facpub/360
Running title: Generalized Species Abundance Distributions

Keywords: species abundance distribution, macroecology, metabolic theory, resource partitioning, body-size, biomass, energy use, size-density relationship, size-energy relationship, body-size distribution

Ideas and Perspectives, 141 words in abstract, 7859 words in manuscript, 87 references, 3 figures, 2 tables, 2 boxes (should be 150 words in abstract, 7500 words in manuscript, 80 references and 10 figures).
Person to whom correspondence should be sent:

Hélène Morlon

Center for Ecology and Evolutionary Biology, 335 Pacific Hall

5289 University of Oregon, Eugene, OR 97403-5289

Tel: (541) 346 5674; Fax: (541) 346 2364; morlon.helene@gmail.com
Taking species abundance distributions beyond individuals

Abstract

The species abundance distribution is one of the few “universal” patterns in ecology. Research on this fundamental distribution has been focused to the study of numerical counts, irrespective of the traits of individuals. Here we show that considering a set of Generalized Species Abundance Distributions (GSADs) that encompass several abundance measures, specifically numerical abundance, biomass and resource use can provide novel insights into organizational forces at play in ecological communities. We first leverage a comprehensive dataset to investigate the similarity and difference between GSADs. We then use classical probability theory and allometric scaling laws to explore theoretical linkages among GSADs and between GSADs and other common macroecological distributions. Finally, we show how examining different GSADs in natural systems can help with assessing determinants of community structure. Broadening species abundance distributions to encompass multiple abundance measures opens fresh perspectives in biodiversity research.

Introduction

The species abundance distribution (SAD) describes the commonness and rarity of species in ecological communities. The SAD is one of the oldest patterns in ecology (Fisher et al., 1943; Preston, 1948; MacArthur, 1957), and remains at the core of ecological research (McGill et al., 2007; Volkov et al., 2007; Alonso et al., 2008; Dornelas and Connolly, 2008). Repeatable patterns in the shape
of the SAD across taxa and environments (the hollow curve, with few dominant and many rare species) suggest that this distribution reflects common underlying principles that structure ecological communities. The SAD has been and is still widely used to test alternative hypotheses on determinants of community structure and biodiversity (Caswell, 1976; Hubbell, 2001; Etienne and Olff, 2005; Chave et al., 2006; McGill et al., 2006b).

Recent work on the SAD has focused on the frequency distribution of number of individuals per species (numerical abundance). This is a natural way to characterize SADs and predictions for this type of abundance distribution are available for a variety of models. The use of numerical abundance has been inspired by population ecology and evolutionary biology, where it is an important determinant of the demography and genetic variability in populations. However, numerical abundance is only one way to characterize the relative abundance of species within communities. Studies inspired by systems ecology or macroecology have measured abundance in terms of the biomass and/or resource use of species populations, to better understand compartmentalization and fluxes of carbon, nutrients and energy in ecosystems (Brown and Maurer, 1986; Maurer and Brown, 1988; Brown and Maurer, 1989; Thibault et al., 2004; Connolly et al., 2005; Harte et al., 2008). Also, instead of numerical abundance, biomass or resource use, field ecologists have sometimes used proxies for these measures, such as percent cover or point quadrat frequency (Tokeshi, 1999; Latimer et al., 2005).

Numerical abundance, biomass, resource use, or proxies for these measures are not necessarily shared equivalently between species in a community. The most numerous species do not always contain the most biomass or use most resources, as they sometimes contain many small individuals. Yet, ecologists have viewed the “classical” SAD (constructed from numerical abundance) to represent the outcome of rules governing how species partition resources (Motonura, 1932; MacArthur, 1957, 1960; Sugihara, 1980; Tokeshi, 1990, 1996; Marquet et al., 2003). The underlying assumption was that individuals across species are comparable in their body-size and resource use, as shown by a quote from MacArthur (1960):

“It is assumed that the species whose abundances are being compared are roughly of same size, so that an individual of one species is comparable to an individual of any of
In cases where all individuals of all species are of the same size and use the same resource, SADs based on the numerical abundance, biomass, or resource use of populations will provide the same information about community structure. This assumption is likely to be violated in natural systems for several reasons. First, body-sizes typically vary by several orders of magnitude between species (Brown, 1995; Clauset and Erwin, 2008; Roy, 2008), and large-bodied species use in general more energy than small-bodied ones (Kleiber, 1947; Peters, 1983). Second, individuals of the same size can have a very different resource use, for example due to different internal nutrient concentrations (Sterner and Elser, 2002). Third, it has been increasingly realized that individuals within populations of the same species can strongly vary in size, for example due to age structure, with important consequences for ecological interactions (Werner and Gilliam, 1984; De Roos and Persson, 2001; Kooijman, 2000). It is thus reasonable to question the equivalence between abundance distributions based on number of individuals, biomass or energy use. Partly for this reason, the discussion on the use of SADs drifted away during the last decades from the perspective of resource partitioning and ecosystem functioning to focus mostly on information (diversity) aspects.

If abundance distributions constructed from alternative measures for the abundance of species are not equivalent, three fundamental questions arise: i) What are the similarities and differences in the way species partition individuals, biomass and resource use? ii) How are these different distributions linked? iii) What insights can be gained from exploring the distribution of biomass and resource use among species? These questions have generated a suite of papers with equivocal conclusions (Harvey, 1987; Sugihara, 1989; Pagel et al., 1991; Taper and Marquet, 1996). In these papers, the authors draw upon classical allometries (Kleiber, 1947; Damuth, 1981; White et al., 2007) to investigate the theoretical link between the distributions of numerical abundance, biomass and resource use. Some of the results suggest that species divide individuals, biomass or resource use in similar ways (Sugihara, 1989; Taper and Marquet, 1996) while others suggest important differences (Harvey, 1987; Tokeshi, 1990; Pagel et al., 1991). Since then, few papers have examined SADs in parallel using alternative measures for abundance (although see Thibault et al., 2004 and Connolly et al., 2005).
In this paper we argue that research on the relative abundance of species will benefit from generalizing the “classical” SAD based on numerical abundance, into a set of Generalized Species Abundance Distributions (GSADs) that encompass alternative measures of abundance, specifically biomass and resource use. We start with an empirical analysis of four major taxonomic groups at various spatial scales and show that specific GSADs differ from one another. We propose that exploring the interrelationship between GSADs can provide insights into the determinants of community structure. In order to link GSADs, we base on previous work using allometric scaling rules, particularly the allometry between population density and body-size, and between body-size and energy requirements (as a proxy for resource use). We show that allometric scaling rules not only link GSADs one to the other but also to two other important macroecological distributions: the species-size distribution (how many species of each body-size) and the species-individual energy distribution (how many species of each individual energy use) (Table 1). We discuss the implications stemming from our analysis for the understanding of community structure and diversity.

**Generalized species abundance distributions are not equivalent**

There is to date no consensus on whether the apportionments of individuals, biomass and resource use across species within a guild are equivalent (Harvey, 1987; Sugihara, 1989; Pagel et al., 1991; Taper and Marquet, 1996). If there is no variation in body-size or if numerical abundance, biomass and resource use are proportional to one another, then patterns observed for different abundance measures should be similar. Theories that ignore interspecific differences or assume that all individuals are ecologically equivalent implicitly or explicitly assume that this is the case (MacArthur, 1960; Hubbell, 2001). However, different abundance measures are not necessarily well correlated (Saint-Germain et al., 2007, see Fig.2) and fundamental differences between SADs constructed from different abundance measures may well occur. Below, we leverage data from a wide variety of taxonomic groups and habitats to test the equivalence of alternative species abundance distributions. Throughout, we use the terminology “generalized species abundance distributions” (GSADs) to refer to frequency distributions constructed from any measure of abundance (individuals, biomass or...
resource use), and the terminology “species individuals distribution” (SID) to refer to the “classical” SAD (the one constructed from numerical abundance). SBD and SED refer to the species biomass and energy distributions, respectively (Table 1).

**Data and Methods**

We use data from a total of 1533 communities spanning four major taxonomic groups: birds, fish, small mammals and trees. The bird data is from the Breeding Bird Survey of North America (BBS; n=1400; Robbins et al. 1986; Sauer et al. 2008), the fish data from a country scale survey of Trinidad (n=76; Magurran and Phillip, 2001), the small mammals data from the Sevilleta LTER (Ernest et al., 2000) and the Portal Project (n=7; Brown, 1998; Ernest et al., submitted) , and the tree data from a 50 ha tropical forest plot at Barro Colorado Island (BCI; n=50; 1ha plots; Condit, 1998; Hubbell et al., 1999, 2005). Each bird species is assigned the characteristic mass found in Dunning (1995). The population biomass of each fish species comes from a direct measurement in the field. The population biomass of each mammal species is obtained from summation of individual body-size measurements. Individual tree biomass is estimated from an empirical interspecific allometry based on diameter at breast height (Brown, 1997), providing a first good approximation of more specific allometries incorporating wood density (Chave et al., 2003). We use energy utilization as a proxy for resource utilization. Empirical data on the energy use of each species population within sampled communities is typically unavailable. We estimate per-capita energy use for birds and mammals based on published allometry on field metabolic rates for the respective groups (Nagy et al., 2001) . For fish and trees, we base upon published allometry on resting metabolic rates (Gillooly et al., 2001). A more detailed description and treatment of the data sources can be found in Appendix S1.

We compare the three different GSADs (frequency distribution of individuals, biomass and energy requirements per species) using three simple characterizations: 1) the variance of log-transformed abundance values (a measure directly related to many common evenness metrics (May, 1975; Smith and Wilson, 1996) and used in previous papers comparing GSADs (Harvey, 1987; Sugiura, 1989; Pagel et al., 1991; Taper and Marquet, 1996)) 2) the skewness of log-transformed
abundance values (e.g., lognormal distributions have no skew) 3) visual evaluation of the variation of the shape of the distributions with spatial scale (i.e. unveiling, Preston, 1962; Nee et al., 1991; Connolly et al., 2005). In combination these three characterizations provide a broad and general picture of the similarities and differences between species abundance distributions based on different measures. In evaluating these results the reader should note that while numerical abundance and biomass are either measured directly or estimated with good confidence, energy use is necessarily calculated by making assumptions about the form of the relationship between metabolic rate and body-size (Appendix S1). Estimation of the energy use of species is based on the best available data, but is still a rough approximation. Results related to energy distributions should thus be interpreted with caution.

**Evenness**

It is well known that the number of individuals is not equitably distributed among species; communities are usually composed of many rare and few common species. The degree of this non-equitability can be characterized by a variety of different measures of evenness (Smith and Wilson, 1996). We quantify evenness simply through a measure of the variance of the logarithm of the abundance measures (more even distributions have smaller variance). For example, a totally even distribution would correspond to a community where all species are equivalent (i.e., no variance). Using the log-transformed data focuses on the shape of the distribution as usually characterized by ecologists (e.g. the famous “Preston plots”, 1962) and eliminates any dependence on units.

Overall the empirical data show that SIDs are more evenly distributed than SBDs and SEDs, and SEDs more evenly distributed than SBDs (Figure 1 a-c and Appendix S2). This pattern occurs across all of the datasets, though it is less consistent in the fish data. These results suggest that, contrary to what has been proposed before (Sugihara, 1989; Taper and Marquet, 1996), the partitioning of individuals, biomass and energy are not equivalent processes. While the datasets behave in a qualitatively similar manner, they clearly differ from one another in the specifics of both evenness values and the relationships between them. The most extreme example is the data on the tropical forest at BCI, which exhibits very even distributions of abundance and extremely
uneven distributions of biomass compared to the other datasets. The variance in any of the log abundance measure typically increases markedly with the variance of any other log abundance measure, though in the fish data these variance-variance relationships are weak. For example, the slope of the $\text{Var} \left[ \log N \right] - \text{Var} \left[ \log M \right]$ relationship (which describes the decrease in evenness in biomass when the evenness in numerical abundance decreases) differs widely between datasets. It is not significantly different from zero for the fish, close to one for the mammals, significantly steeper than one for the birds, and much steeper than in any other dataset for the plants.

**Skewness**

Skewness is a measure of asymmetry in a frequency distribution based on the third central moment, widely used to characterize SADs (e.g. Hubbell, 2001; Magurran, 2003; McGill, 2003). Differences between the skewness of GSADs based on different measures of abundance are less consistent among datasets than the differences in evenness (Figure 1 d-f and Appendix S2). The mammal communities appear to have equivalent skewness regardless of the abundance measure. The other three datasets do tend to exhibit changes in the average sign/magnitude of the skewness among the different abundance measures. The bird data does not show a clear tendency for the skew of the SID (about as much positively versus negatively skewed distributions), while the SBD and SED are generally positively skewed. The fish data show the opposite pattern with more communities with positively skewed SIDs, but no clear tendency for the SBD and SED. The plants show the most extreme differences, clearly shifting from positively skewed SIDs to negatively skewed SBDs and SEDs. In sum, with the exception of the mammals, all datasets agree with the trend that skewness differs with the measure chosen for abundance. Some of the observed changes in sign, notably the change in the plant dataset, may be related to the results on unveiling discussed below.

**Unveiling**

The “veil effect” is a well-known phenomenon by which SADs appear truncated if the rarest species are not sampled (Preston, 1962). The veil effect disappears as sampling effort increases, for example as the spatial extent of the sample increases. In a recent paper, Connolly et al. (2005) showed that
the species individual distribution unveils at much larger spatial scales than the species biomass distribution.

Our data suggests that what Connolly et al. referred to as a “scale discrepancy”, a phenomenon in which the species individual and the species biomass distribution unveil at different spatial scales, is not specific to the coral systems they studied as it also occurs in the bird, fish, and tree datasets that we analyze (Figure 2). This phenomenon appears to be a general pattern: it applies to all the systems with data available at multiple scales of observation, and also holds for the species energy distribution (though in the case of the fish data the unveiling is less rapid for energy than for biomass). Further research is needed to fully understand why the numerical abundance unveils more slowly than the other measures of abundance, but this may result in part from the fact that biomass and energy use are continuous measures, whereas numerical abundance is inherently discrete (EW, personal observation).

The difference in the scale at which different GSADs unveil may be responsible for some of the observed differences in the skewness of distributions. In particular the shift in the skewness of the plant GSADs from positive using numerical abundance to negative using energy/biomass likely reflects the fact that at 1 ha the SID is far from unveiled and is thus right-skewed, whereas the energy and biomass distributions are fully unveiled thus revealing their long left tail.

Conversion framework

The empirical patterns of evenness, skewness and unveiling presented above clearly demonstrate that the distributions of individuals, biomass and energy (GSADs) are not equivalent. However these GSADs are tightly linked through allometric relationships between number of individuals, body-size, and individual energy use. These allometries not only link GSADs one to the other, but also to other macroecological distributions, such as the frequency distributions of body-size (the well-known species-size distribution SsD; Hutchinson and MacArthur, 1959), and individual energy use (Figure 3 and Appendix S3). Below we outline the mathematical framework that allows for linking all form of these macroecological distributions. We aim for generality, and investigate the
conversion between any two diversity distributions $S_X$ and $S_Y$ constructed from any measurements $X$ and $Y$. To render the formulas more concrete, $X$ and $Y$ may be replaced by number of individuals, biomass, species resource use, body-size or per-capita resource use (e.g. Box 1).

**General conversion framework**

The conversion between two diversity distributions $S_X$ and $S_Y$ is given by basic probability theory:

$$S_Y(Y) = \int P(Y|X) S_X(X) \, dX$$

(1)

$P(Y|X)$ is the probability that a species with measurement $X$ also has measurement $Y$ (Appendix S3). In this paper for simplicity we use the integral notation for both continuous and discrete $X$. The integral should be interpreted as a sum when $X$ is discrete. If there exists a functional relationship between $Y$ and $X$ (e.g. a species’ numerical abundance is in part explained by the species body-size, or the species biomass), we can write:

$$Y = f(X) + Z$$

(2)

where $Z$ is a random variable independent of $X$ (e.g. a species’ numerical abundance is also explained in part by other traits than body-size, or other factors than the species biomass).

In the particular case where $X$ is the main explanatory variable for $Y$, $Z$ can be thought of as a residual error $\varepsilon$ independent of $X$, usually normally distributed. Denoting $P_\varepsilon$ the probability density distribution of this error, it is straightforward to show (Appendix S3) that the distribution $S_Y$ is related to the distribution $S_X$ by:

$$S_Y(Y) = \int P_\varepsilon(Y - f(X)) S_X(X) \, dX$$

(3)

This general equation shows that any two diversity distributions $S_X$ and $S_Y$ stem one from another provided a functional link between $X$ and $Y$. The equation can be used to derive results for any specific form of the distributions of $X$ and $Y$, the functional relationship between them,
and the form of the residual error (Table 2). By substituting two of \( N \), \( M \) or \( E \) in \( Y \) and \( X \), the formulas in Table 2 show that there is in general no reason to expect the shape of the species individuals, biomass and energy distributions to be same. Below we examine the case when the functional relationship between measurements can be considered power-law.

**Case of power-law allometries**

Power-laws allometries, which take of the form \( cz^a \), constitute convenient approximations that render derivations analytically tractable. Power-laws are certainly not universal, but there is empirical evidence for their existence in nature. For example, empirical and theoretical work support that population density scales with body-size as \( N \propto \overline{m}^{-\frac{3}{4}} \) (size-density relationship, Damuth, 1981, 1987; Peters, 1983; White et al., 2007) and individual energy use with body-size as \( e \propto m^{\frac{3}{4}} \) (size energy relationship, Kleiber, 1947; West et al., 1997). Noting that the biomass of a species is given by \( M = N\overline{m} \) and its’ energy use by \( E = N\overline{e} \), it is straightforward that if the size-density relationship is power-law (let note its exponent \( a_{N|\overline{m}} \)) and the size-energy relationship is power-law (exponent \( a_{e|m} \)), then relationships between all pairwise comparisons of \( N \), \( M \), \( E \), \( \overline{m} \) and \( \overline{e} \) are power-law (ignoring intraspecific size variation, we assume that the relationship between \( \overline{e} \) and \( \overline{m} \) is power-law).

For example, \( a_{M|N} = 1 + \frac{1}{a_{N|\overline{m}}} \) and \( a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|\overline{m}}} \) (Appendix S3).

We consider here log-transformed data and assume a power-law allometry between \( X \) and \( Y \), with exponent \( a_{Y|X} \):

\[
Y = \log (c_{Y|X}) + a_{Y|X} X + Z_{Y|X} \tag{4}
\]

The conversion between \( S_X \) and \( S_Y \) is directly given by adequate substitution in Equation 3 (Table 2 and Appendix S3). The formulas presented in Table 2 can be used to derive the full shape of distributions. For example, Loehle (2006) showed that the geometric species individuals distribution can result from a uniform species-size distribution and an exact power-law size-density relationship. This result is easily reproduced using our framework (with \( Y = N \) and \( X = \overline{m} \); Appendix S3). Our formulas (Table 2) more generally demonstrate that in principle, any species
individual distribution could be derived from a specific species-size distribution and size-density relationship.

Many studies of macroecological distributions (GSADs, SsD or SeD) are more interested in simple characterizations of the distributions based on their higher moments, such as the evenness and skewness, than in their shape. Here, we write the general relationship between the evenness of two distributions constructed from any two measurements \( X \) and \( Y \). In Box 1, we write the more concrete case of the relationship between the evenness of the species individuals and biomass distributions. Appendix S4 presents the case of the species individuals and energy distributions.

Following our empirical analyses, we quantify evenness through a measure of the variance (more even distributions have smaller variance). There are two ways to describe the functional relationships between \( X \) and \( Y \): 1) \( Y \) is functionally determined by \( X \) (Equation 4), with the underlying idea that “\( Y \) stems from \( X \)”. For example, the biomass of a population may be largely a consequence of how numerous the population is if processes act primarily to determine the number of individuals in a population (e.g. processes of birth, death, speciation and migration) 2) \( X \) is functionally determined by \( Y \), with the underlying idea that “\( X \) stems from \( Y \)”. For example the density of a population may be largely a consequence of its biomass if processes act primarily to determine population biomass (e.g. resource allocation, spatial disturbances). These two different ways to write the functional relationship between \( X \) and \( Y \) lead to two different relationships between the variance of \( X \) and \( Y \) (Taper and Marquet, 1996).

In the first case (\( Y \) functionally determined by \( X \)),

\[
Y = \log \left( c_{Y|X} \right) + a_{Y|X}X + Z_{Y|X}
\]

with \( Z_{Y|X} \) independent of \( X \). In this case,

\[
Var\left[ Y \right] = a_{Y|X}^2 \cdot Var\left[ X \right] + Var\left[ Z_{Y|X} \right]
\]

In the second case (\( X \) functionally determined by \( Y \))
\[ X = \log(c') + \frac{1}{a_{X|Y}} Y + Z_{X|Y} \]  

(7)

with \( Z_{X|Y} \) independent of \( Y \). In this case, \( \text{Var}[X] = \frac{1}{a_{Y|X}^2} \text{Var}[Y] + \text{Var}[Z_{X|Y}] \) and

\[ \text{Var}[Y] = a_{X|Y}^2 \text{Var}[X] - a_{Y|X}^2 \text{Var}[Z_{X|Y}] \]  

(8)

Note that here the variance of \( X \) and \( Y \) are not (measurement) errors of a single value of \( X \) and \( Y \) for a single species. Rather, \( \text{Var}[X] \) and \( \text{Var}[Y] \) pertain to the overall variance in observed \( X \) and \( Y \) values across species. Equations 6 and 8 show that the relationship between the variance of two macroecological distributions \( S_X \) and \( S_Y \) depends on the steepness of the allometric relationship between \( X \) and \( Y \) and on the model of functional relationship between \( X \) and \( Y \) ("\( Y \) stems from \( X \)" or "\( X \) stems from \( Y \)"). Substituting \( X \) and \( Y \) in these equations by \( N, M \) and \( E \), with \( a_{M|N} \) and \( a_{E|N} \) as given above, we deduce that the variance in biomass increases faster than the variance in numerical abundance for a size-density slope shallower than \(-\frac{1}{2}\) (Box 1, Appendix S4) and the variance in energy use increases faster than the variance in numerical abundance for a ratio size-density exponent/size-energy exponent greater than \(-\frac{1}{2}\) (Appendix S4). We can further examine under which conditions biomass and energy are more or less equitably distributed than individuals (Box 1, Appendix S4). Also, Equations 6 and 8 may be used to understand relevant "causal models" in natural communities (Taper and Marquet, 1996). Indeed, variance values and allometric exponents can be measured from field data. Substituting in equations 6 and 8, only one equation can hold. The "correct" equation should correspond to the "causal model" at play in the community (Box 1). Below, we explore the relationship between the skewness of macroecological distributions, and show that skewness might give another means to understand processes at play in ecological communities.

Consider the case where \( Y \) is functionally determined by \( X \) (Equation 4). The skewness of a variable \( X \) is given by:

\[ \text{Skew}[X] = \frac{\text{E}[(X - \text{E}[X])^3]}{(\text{E}[(X - \text{E}[X])^2])^{\frac{3}{2}}} \]  

(9)
We show that for any centered and un-skewed random variable $Z_{Y|X}$ (e.g. normally distributed on a log-scale; Appendix S3):

$$Skew [Y] = \frac{a_{Y|X}^3 E[(X - E[X])^3]}{(a_{Y|X}^2 Var[X] + Var[Z_{Y|X}])^{\frac{3}{2}}} \tag{10}$$

The denominator in Equation 10 is always positive, so that the sign of the skewness is given by the sign of the numerator, i.e. it is conserved for $a_{Y|X} > 0$, and inversed for $a_{Y|X} < 0$. When applied to population density $Y = N$ and body-size $X = m$, this result suggests that a decreasing size-density relationship should yield to species individuals and species-size distributions with opposite skew (i.e a left-skewed species individuals distribution should correspond to a right-skewed species-size distributon and vice-versa, Figure 3). Substituting now $X$ and $Y$ by $N$, $M$ and $E$, with $a_{M|N}$ and $a_{E|N}$ as given above, we deduce that the species individuals and biomass distributions should have opposite skew for $0 > a_{N|m} > -1$ and that the species individuals and energy distributions should have opposite skew for $0 > a_{N|m} > -a_{e|m}$ (Appendix S3). The absolute value of the skewness is unchanged for a perfect allometric relationship between $X$ and $Y$ ($Z_{Y|X} = 0$). In the more realistic case when the allometry is not perfect, the absolute value of the skewness of the “explanatory” variable ($X$) is always larger than that of the “explained” variable ($Y$). Hence, comparing the skewness of macroecological distributions may help understand “causal models” underlying community structure. With the assumption that $Z_{Y|X}$ is un-skewed, Equation 10 predicts that if $X$ is un-skewed, then $Y$ is also un-skewed. In Appendix S3, we generalize Equation 10 to the case when $Z_{Y|X}$ is skewed, which is likely to be the case when the power-law assumption is not valid (e.g. there is some evidence that the species-size distribution might be unimodal, not power-law; Brown and Maurer, 1989; Marquet et al., 1995). We show that in this case, $Y$ is also expected to be skewed. Below, we discuss the ecological implications of our results.

**Ecological implications**

Our empirical and theoretical analyses of the links between species abundance distributions constructed from different abundance measures (GSADs) demonstrate that they are not equivalent.
Although GSADs differ one from another, they can be interrelated through allometric scaling laws. They are also related to other fundamental patterns including the species-size and species individual-energy distributions. All of these macroecological distributions (the GSADS, SsD and SeD) relate one to the other through the size-density and size-energy relationships. Their interrelationship also depends on the “causal” model at play in communities. We discuss these results in light of previous work, and show how they can be used to help understand the forces that organize communities.

**Species individuals, biomass and energy distributions**

There has been widespread belief in ecology that species individuals, biomass and resource use distributions are equivalent, and this may be one of the reasons why the species biomass and resource use distributions are under-represented in the literature. The validity of the equivalence between the different abundance distributions was challenged over twenty years ago (Harvey, 1987), generating a 10-year long debate (Sugihara, 1989; Pagel et al., 1991; Taper and Marquet, 1996) that ended in the following terms (Taper and Marquet, 1996):

“The apportionment of individuals and resources, among species, are equivalent processes”.

Our results contradict this statement. Empirical data across a wide variety of taxonomic groups show that species abundance distributions constructed from different measures of abundance (GSADs) are not equivalent. These results are supported by theoretical arguments. Below we discuss the cause and consequence of differences in the shape, evenness, skewness, and unveiling of GSADs.

**Shape**

The shape of SADs has long fascinated ecologists (Fisher et al., 1943; Preston, 1948). Understanding the specifics of the “hollow curve” has been central to the development of ecological theory, and remains at the core of ecological research (McGill et al., 2007; Volkov et al., 2007; Alonso et al., 2008; Dornelas and Connolly, 2008). Our results (Table 2) suggest that the shape of SADs is not generally conserved when converting from one abundance measure to another. The fact that
GSAD shape is not conserved suggests that a theoretical prediction for one abundance measure is not directly applicable to another. For example, theory based on resource partitioning assumptions (e.g. (MacArthur, 1957; Sugihara, 1980)) will only directly predict the shape of species energy distributions, and theory based on population dynamics (e.g. the Neutral Theory of Biodiversity; Hubbell 2001) will only directly predict the shape of species individuals distributions. In the extent to which abundance is related to body-size or energy use, the conversion from one GSAD to another may be approached using the general formulas presented in Table 2. Such conversions will extend the breadth of current ecological theory from one to multiple GSADs. From a practical standpoint, such conversions will allow for assessing the shape of GSADs that are difficult to measure empirically. For example, our conversion framework provides a means for predicting species-energy distributions, which are relevant to resource-based models but often empirically intractable. This conversion framework could also provide a tool for estimating the classical SAD in systems where this is currently challenging, for example in microbial systems (Quince et al., 2008).

Evenness

The evenness, or variance of GSADs has both empirical and theoretical relevance. The variance of GSADs reflects the equitability with which species partition individuals, biomass or resources, and is crucial to our understanding of ecological organization in natural communities (Harvey, 1987; Sugihara, 1989; Pagel et al., 1991; Taper and Marquet, 1996). The variance played a central role in the early theoretical developments of SADs. Variance is a key parameter in Preston’s celebrated “canonical” lognormal distribution (Preston, 1962). Under the canonical hypothesis, the variance of the species abundance distribution of an ecological community is sufficient to entirely characterize the distribution, and therefore the number of individuals and the number of species in that community. Understanding the mechanisms underpinning the canonical lognormal has been a focus in ecology (e.g. May, 1975).

Sugihara’s arguments that the sequential resource-division model generates a canonical lognormal are founded on the ability of this model to generate the specific constraint between the SAD variance and species richness characteristic of the canonical lognormal. An implicit component of
Sugihara’s sequential resource-division model is the assumption that the variances of GSADs are equivalent. Sugihara’s model is based on niche apportionment, whereas the variance of Preston’s canonical lognormal relates to numerical abundance. Sugihara states: “a niche translates ultimately into numbers of organisms (or biomass)” (Sugihara, 1980). Sugihara’s niche apportionment model only generates a canonical lognormal if variances in numerical abundance, biomass and resource-use are equal, which has been highly debated in the literature (Harvey, 1987; Sugihara, 1989; Pagel et al., 1991; Taper and Marquet, 1996). Taper and Marquet’s work (1996) brought momentary closure to this debate by legitimizing the assumption of equivalence between variances of different abundance measures.

Our empirical results disagree with the conclusions of Taper and Marquet (1996), and instead suggest that different GSADs vary widely in their evenness (Figure 1). We find that individuals are the most equitably distributed, followed by resource-use, and finally biomass, which is the most un-equitably distributed. Our results differ from Taper and Marquet due to an important difference in our approach to analyzing empirical data. Taper and Marquet estimate the empirical variance in energy among species using a theoretical relationship linking variances in energy and numerical abundance. As the authors rightly note, their estimation likely suffers from error propagation. In our empirical analyses, we directly estimate the energy requirement of each species from fitted relationships between field or resting metabolic rate and body-size, and calculate the variance among species a posteriori. Because we have data on body-size (but for the birds), we eliminate the principal source of error in the theory-based variance prediction used by Taper and Marquet. One limitation of our approach is that we ignore the error introduced by using the size-energy allometry. However, including error propagation would increase the variance in energy-use, given the biologically realistic assumption that metabolic rate is determined by body-size (“body-size causal” model). This suggests that the disparity in variances between the species energy distribution and the species individuals distribution is likely to be even greater than we observed. Even though our estimation of the variance in energy use is more direct than that of Taper and Marquet, it is still only estimation. Yet our analyses of the variance in biomass and numerical abundance directly stem from field measurements. Thus our results unequivocally show that GSADs are not equivalent.
We supplement our empirical comparisons of variance across GSADs by leveraging theoretical relationships between them. Our approach is similar to that of Pagel et al. (1991), who examine the theoretical relationships between variances for the three different abundance measures. From these relationships, they deduce the allometric scaling exponents relating density, size and energy-use that would achieve the equality of variances. While deriving their variance-variance relationships, Pagel et al. (1991) implicitly assume numerical abundance to be the “causal variable”, which is not necessarily ecologically realistic (Taper and Marquet, 1996). We re-write these relationships under a “N causal”, “M causal” and “E causal” model (Box 1 and Appendix S4). From these relationships, we can deduce the conditions necessary for local communities to cluster around a 1:1 line on variance-variance graphs (thus leading to the equality of variances). We find that the equality of variances across local communities is only expected to be achieved for very specific values of the size-density and size-energy allometric exponents (Box 1 and Appendix S4), values that do not hold in any of the dataset we analyze.

Our theoretical approach to comparing variance-variance relationships can be used to understand the potential causes of several patterns in our empirical data (Figure 1). A first notable pattern is that communities in all datasets and in both the $\text{Var}[M] - \text{Var}[N]$ and $\text{Var}[E] - \text{Var}[N]$ graphs cluster around a well defined line, suggesting that there exists an underlying size-density and size-energy allometry common to all local communities. A second notable pattern is that the line around which communities cluster is steeper than and lies above the 1:1 line. The steepness of the relationship may be explained by a size-density relationship exponent less than $-\frac{1}{2}$. The steepness of the $\text{Var}[E] - \text{Var}[N]$ relationship may be explained by a ratio of the size-density to the size-energy exponents shallower than $-\frac{1}{2}$. Intuitively, a significant fraction of the most numerous species are big, account for a large part of community biomass, and use most resources, while a significant fraction of the rare species are small, account for only a small part of community biomass, and get a small share of resources. These tendencies are amplified as the variation in numerical abundance between species increases.

A third notable pattern is that the specifics of the tendencies described above, in particular the steepness of the variance-variance relationships, vary among taxa. These differences stem from
differences in the size-density and size-energy allometries. The shallower the size-density allometry, and the smaller the absolute value of the ratio between size-energy and size-density exponents, the steeper the variance-variance relationships around which communities cluster. The steepest variance-variance relationship is for the trees. Trees are indeterminate growers, with a particularly weak dependency of population density with body-size. Also, a large fraction of aboveground biomass is stored in large trees in tropical forests (Chave et al., 2003). When a species is numerous in a local community, the probability of encountering large trees increases, giving a possible explanation for the steep increase in the variance in biomass with increasing variance in numerical abundance.

The conclusion stemming from our analyses on evenness, that biomass and resources are less equitably distributed among species than individuals, goes against the idea embedded in the so-called “energy equivalence rule” (Damuth, 1981). Nearly thirty years ago, Damuth (1981) gathered data showing that the numerical abundance of a species decreases as a $-\frac{3}{4}$ power of its body-size. Combining this observation with the scaling of metabolic rate with body-size ($e \propto m^{\frac{3}{4}}$) yields:

$$E \propto N \sigma \propto m^{-\frac{3}{4}}m^{\frac{3}{4}} \propto m^0$$

which suggests that the resource use of a species is independent on population density and body-size. The energy equivalence rule is attractive because it suggests that physical constraints and ecological processes balance in a sort that small-bodied and large-bodied species receive an equal amount of resources (White et al., 2007). Under this rule, one would expect the variance in energy among species to be approximately constant with respect to (and smaller than) the variance in numerical abundance. The energy equivalence rule is clearly not achieved in our datasets, where variance in energy and numerical abundance are highly correlated, and variance in energy is not smaller, but larger than variance in numerical abundance. The ratio between the size-density and the size-energy exponent should be less than $-\frac{1}{2}$ for resource-use to be more equitably distributed than individuals, and equals to $-1$ for the energy equivalence rule to hold. Instead, this ratio is greater than $-\frac{1}{2}$ in our data, leading resources to be less evenly distributed than individuals.

The fact that biomass and resources are less equitably distributed among species than individuals has important ecological and evolutionary consequences. In the four taxonomic groups analyzed,
small-bodied species leave below the density for which they would receive an equal share of the resources, while large-bodied species leave above this density. A first possible explanation for the disproportionately small share of resources utilized by small species involves the patchy nature of resources perceived and consumed by small species, confining them to small islands within the larger community. A second explanation stems from the idea that small species are more sensitive to density-independent sources of mortality than large species. Hence, large species may be leaving at the carrying capacity imposed by resource availability, while small species would be leaving below this carrying capacity (Brown and Maurer, 1986; Maurer and Brown, 1988; Brown and Maurer, 1989; Pagel et al., 1991). Large species have an advantage in resource monopolization, which supports the idea that “bigger is better”. It has been hypothesized that the selective advantage of large-bodied species in resource monopolization is responsible for the observed increase in body-size in the course of evolution (Cope’s rule; Brown and Maurer, 1986).

**Skewness**

Investigating the skewness of GSADs in addition to their evenness is central to the theory on SADs. Research into the processes underlying the species-abundance distribution has been highly influenced by observations on its skewness. Early statistical explanations for the origin of the shape of the SAD May (1975) were largely based upon the normal (un-skewed) nature of the SAD (plotted on a log-scale). More recent developments have been motivated by the repeated observation that SADs tend to exhibit a significant left skew (Magurran, 2003; McGill, 2003). This observation has in particular been one of the major justifications for the Neutral Theory of Biodiversity (Hubbell, 2001). If the left skew is not consistent across GSADs, then this justification for the Neutral Theory will only be valid for the measures of abundance where the left skew occurs. Our empirical and theoretical results suggest that the skewness (Figure 1 and Equation 10) between species individuals, biomass and energy distributions are not expected to be the same. For example, an un-skewed species individuals distribution (e.g. the lognormal) can produce a skewed species energy distribution if the variation around power-law allometries is skewed (Appendix S3). Also, the direction of the skewness is not necessarily conserved between abundance measures. Under
the assumption of un-skewed error around allometric scaling laws, species individuals and biomass distributions are expected to have opposite skew as soon as the slope of the size-density relationship is shallower than $-1$, which is usually the case in natural communities. Also, species individuals and energy distributions are expected to have opposite skew as soon as the slope of the size-density distribution is shallower than the slope of the size-energy distribution, which is the case in our data and is also supported by others (Brown and Maurer, 1986, 1987).

**Unveiling**

An other important difference between GSADs stems from the scale at which they unveil: biomass and energy distributions unveil at a much smaller spatial scale than species individuals distributions. These results are consistent with patterns observed by Connolly et al. (2005) in coral and fish communities. Basing upon Taper and Marquet (1996), Connolly et al. (2005) suggest that the non-equivalence of the species individuals and biomass distributions (because they unveil at different scales) provides support against niche-based origins of the species abundance distribution. Our results suggest that species individuals and biomass distributions are not expected to be similar, irrespective of which ecological processes generate them. The difference in the spatial scale at which species individuals and biomass distributions unveil appears to be a general phenomenon, and also applies to the energy distribution. Further research will be necessary to fully understand why species biomass and energy distributions unveil faster than species individuals distribution, but preliminary results suggest that the difference might arise from their continuous rather than discrete nature (Ethan White, personal observations). The rapid unveiling of biomass and resource use suggests that GSADs based on these measures may avoid a number of problems associated with the distribution of numerical abundance. The number of individuals or areas sampled in typical ecological surveys are often too little to completely unveil species individuals distributions, particularly in species rich systems, such as tropical or microbial communities (Connolly et al., 2005; Quince et al., 2008). Reconstructing the shape of the “true”, “un-truncated” species individuals distribution is not an easy task (Green and Plotkin, 2007). Estimating the underlying parameters of truncated distributions is also problematic. Focusing on biomass or energy-based GSADs may thus
provide a more consistent and more direct characterization of community structure than individual-based ones.

**Links with other macroecological distributions**

The probability theory we use is very general and offers perspectives for linking frequency distributions in ecology. Frequency distributions are widespread in ecological research, including the “classical” species abundance distribution (the one constructed from individual counts, SID), the species biomass and energy distributions, the species-size and individual-energy distributions, and also frequency distributions of trait values (Reich et al., 2003; Chave et al., 2006). Traits are characteristics of a species that influence its fitness, and therefore its abundance. Body-size and individual-energy use are obvious examples of organismal traits, but many other functional traits are studied in ecology. Ecologists look at species trait distributions (Reich et al., 2003; Chave et al., 2006), trait-trait allometries (Wright et al., 2004) and abundance-trait relationships (McGill et al., 2006a; Shipley et al., 2006). The probability theory we use applies for linking any two GSADs, any GSAD to any species trait distribution (given a mathematical link between traits and abundance), or any two mathematically related traits (Table 2; Stegen and White, 2008). These relationships add to the growing literature on integrating patterns in ecology (Reuman et al., 2008; Southwood et al., 2006).

Linking the “classical” species abundance distribution (SID) to other important macroecological distributions is fundamental for the quest towards “unified” theories that can explain many macroecological patterns in concert (Hubbell, 2001; McGill et al., 2007). The SID is central in this context, because several patterns stem from it; it is even sometimes considered a “master pattern” (McGill et al., 2007). However, most research concerned with linking the SID to other macroecological patterns has concerned spatial biodiversity patterns (Preston, 1962; He and Legendre, 2002; Green and Ostling, 2003; Plotkin and Muller-Landau, 2002; Morlon et al., 2008). Here we formalize the mathematical links between the SID and many other fundamental macroecological patterns including the species biomass and energy distributions, the species-size and individual-energy distributions, and the size-density and size-energy relationships.
Realizing that the species individuals and species size distribution are linked opens a crucial question: which of the two patterns should we try to find mechanisms/models for? The SsD has been and is still the subject of much attention in the literature, and many models have been developed to explain its shape or its skewness (Hutchinson and MacArthur, 1959; Brown and Maurer, 1989; Etienne and Olff, 2004; Allen et al., 2006; Clauset and Erwin, 2008; Roy, 2008). Our results show (as noted before by Loehle (2006)) that the species-individuals distribution (SID) and the species-size distribution (SsD) stem one from another through the size-density relationship. Should we then try to explain the SID, the SsD, or the size-density relationship? We find that the skew of the SID is the opposite of the skew of the SsD provided a negative slope for the size-density relationship. Should we then try to find models that produce a specific skew for the species-individual distribution, or for the species-size distribution? Answers to these questions have potentially far-reaching impacts, because while it is well recognized that body-size is tightly correlated with population density (White et al., 2007), to date the SID and SsD have been studied in isolation. Applying the general formulas provided in our paper with specific shapes for the species individuals or size distributions will help bridge this gap.

Realizing that GSADs are interrelated opens a similar question: which of the SID, SBD or SED should we try to find mechanisms/models for? It is not clear which one of the species individuals, biomass or energy distributions is the closest to processes. If processes of community assembly are dominated by population dynamics, the species individuals distribution should directly reflect these processes, generating an “N causal” model. Instead, if processes of community assembly are dominated by trophic interactions or disturbances acting on population biomass, the species biomass distribution should directly reflect these processes, generating an “M causal” model. Finally, if processes of community assembly are dominated by resource division rules, the species energy distribution should directly reflect these processes, generating an “E causal” model.

Analyzing different GSADs and species-trait distributions simultaneously may help draw inference on organizational forces at play in communities. A difficulty inherent to studying complex ecological systems is that patterns are interrelated (Box 2). Leveraging the imprint of the “causal” pathway on the relationships between frequency distributions may yield insight into the patterns
that are the closest to processes. Taper and Marquet (1996) were pioneers in realizing that the relationships between the variance of GSADs critically depends on the “model of causality”. Our findings support this result, extend its scope to any frequency distribution, and supplement the variance analysis with a parallel result on skewness. Our analyses show that the propagation of error about causal relationships inflates the variance of “non causal” frequency distributions, and reduces their skewness.

The fact that the both the variance-variance and skewness-skewness relationships depend on causality can be used to disentangle “causal” and “non causal” ecological variables. For example, the observed functional relationship between the variance of \( \log(M) \) and the variance of \( \log(N) \) reveals the mechanistic linkages between fundamental allometric constraints within different taxa (Box 1). A positive intercept indicates that population dynamics drive the processes of community assembly. In this case, the main effect of body-size is to constrain those attributes related to population dynamics (e.g. birth and death rates, age at first reproduction). A negative intercept indicates that processes act on the biomass rather than numerical abundance of species. The main effect of body-size is to affect a species’ ability to accumulate biomass, rather than to influence population density via the production or recruitment of new offspring. Our results suggest that taxa differ in the relative importance of processes acting either on population density or population biomass. In birds, the intercept for the relationship between the two variances is positive. As birds are heterotrophic and have determinant growth, body-size likely influences life history attributes related to population dynamics. In trees the intercept is negative. Plants have indeterminate growth, and their dynamics are not likely limited by their body-size. It is possible that the primary influence of a plant species’ body-size is to affect processes that act on biomass, such as disturbance.

**Conclusion**

Recent research on the species abundance distribution has focused on “abundance” measured in terms of individual counts, and has improved the understanding of how individuals are partitioned among species. Our study shows that the partition of individuals, biomass and resource use are
not equivalent process. Our analyses suggest that investigating simultaneously species abundance distributions constructed from different abundance measures will help to decipher major determinants of community structure. Integrating multiple abundance measures in species abundance distribution research will demultiply perspectives on community structure and likely provide strong inference on ecological processes.

Captions

- Table 1: *Measurements, their corresponding distributions and symbol notations*

We denote abundance measures by upper case symbols (e.g. \( N, M \) and \( E \)) and measures at the individual level (per-capita measurements) by lower case symbols (e.g. \( m \) and \( e \)). Overlines stand for averages over individuals of a given species (e.g. \( \overline{m} \) and \( \overline{e} \)). \( S_N(N) \) is the number of species with number of individuals \( N \). \( S_M(M)dM \) is the number of species with biomass between \( M \) and \( M + dM \), \( S_E(E)dE \) the number of species with energy requirement between \( E \) and \( E + dE \), \( S_m(m)dm \) the number of species with size between \( m \) and \( m + dm \) and \( S_e(\overline{e})d\overline{e} \) the number of species with per capita energy requirement between \( \overline{e} \) and \( \overline{e} + d\overline{e} \). The "classical" species abundance distribution (the one constructed from the number of individuals) is coined "species individual distribution" (SID), while the term "generalized species abundance distributions" (GSADs) is used to refer to any of the distributions of numerical abundance (SID), biomass (SBD) or energy use (SED).

- Table 2: *Conversion formulas between macroecological distributions for specific case of allometry and error*

Formulas providing the general conversions from any macroecological distribution \( S_X \) (constructed from measurement \( X \)) to any macroecological distribution \( S_Y \) (constructed from measurement \( Y \)), assuming that \( X \) explains \( Y \) through the relationship \( f \). \( X \) and \( Y \) are typically log-transformed data. On a log-log scale, a power law allometry is linear \( (f(x) = \log(c) + ax) \) and a log-normally distributed error is given by \( P_e(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \).
• Figure 1: *Comparison of evenness (a-c: left panels) and skewness (d-f: right panels) between diversity distributions constructed from numerical abundance (N), biomass (M) and energy use (E)*

Each dot represents a local community. Black: birds; \( n = 1400 \). Blue: fish; \( n = 76 \). Red: mammals; \( n = 7 \). Green: trees; \( n = 50 \). Note that graphs involving energy (\( b, c, e \) and \( f \), grey shade) should be interpreted with caution. Indeed, while numerical abundance and biomass stems from measurements in the field (\( a \) and \( d \)), energy use is estimated using assumptions about the form of the relationship between metabolic rate and body-size (Appendix S1). Evenness is quantified using the variance of log-transformed variables (evenness decreases with increasing variance). Skewness is quantified using \( \text{Skew}[X] = \frac{E[(X-E[X])^3]}{(E[(X-E[X])^2])^{3/2}}. \) Overall, these graphs show that distributions constructed from different measures for abundance are not equivalent.

• Figure 2: *Comparison of unveiling patterns between GSADs constructed from numerical abundance, biomass and energy use*

Plain lines represent GSADs at the largest available spatial scale. Birds: data at the continental scale; Fish: data at the country scale; Trees: data at the 50 ha scale. The mammal data contain only local communities and can thus not be used for this unveiling analysis. Dotted lines represent data averaged over the local communities. Variation around the mean is small and does not influence the general conclusion stemming from these analyses: the species biomass and energy distributions unveil at much smaller scale than the species individuals distribution in all three datasets.

• Figure 3: *Macroecological space for a hypothetical ecological community*

This figure illustrates how the “classical” species abundance distribution (constructed from individual counts, SID) relates to species biomass (SBD) and energy (SED) distributions (a), but also to the species size (SsD) and individual energy use (SeD) distributions (b). Each point represents a species in a hypothetical community. The data are generated assuming: 1) a log-series species individual distribution with parameter 0.9 2) a power-law allometry between the number of individuals and size with exponent \(-\frac{3}{4}\) (Damuth’s exponent) and normally distributed error with variance 0.5
2) a power-law allometry between individual mass and individual metabolic rate with exponent \( \frac{3}{4} \) (Kleiber’s exponent) and normally distributed error (variance 0.5); all individuals are assumed of the same size, so that \( m = \bar{m} \) and \( e = \bar{e} \). The gray shadows in the figure are projections of the species on the 2-dimensional surfaces, showing allometric relationships between any two variables. For example, the projection on the log \( (N) - \log (\bar{m}) \) plane (panel b) is the size-density relationship, and the projection on the log \( (\bar{e}) - \log (\bar{m}) \) plane is the size-energy relationship. The red dots are projections of the dataset on the 1-dimensional axes. The density of these dots constitutes the macroecological distributions (SID, SBD, SED, SsD and SeD), shown in inserts in their familiar histogram form. The energy equivalence rule resulting from the choice of the exponents (same steepness for the size-density and size-energy relationships) is illustrated by the projection on the log \( (E) - \log (N) \) surface (panel a): \( E \) does not depend on \( N \). Note that the right-skewed log-series SID corresponds to a left-skewed SsD (panel b).

**Acknowledgments**

This manuscript is the product of a working group on “Tools and fresh approaches for species abundance distributions” sponsored by the National Center for Ecological Analysis and Synthesis and led by Brian McGill. We are grateful to Brian McGill for organizing the working group and providing his scientific support on the project. EPW led the sub-working group resulting in this paper. The authors recognize that contributing to this paper was one of the key occupations of RSE while he was a Courtesy Research Associate with JLG (and HM) at the University of Oregon. We gratefully acknowledge the funding and logistical support provided by NCEAS. Various individual authors also acknowledge support from NSF, NCEAS, NWO-ALW. We are thankful to Jamie Gillooly for providing us with his data on resting metabolic rates. We also acknowledge the many volunteers as well as USGS and CWS researchers for collecting and making available the Breeding Bird Survey data. ANNE M: Fish data. ETHAN W: Mammals data. The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-
9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Celera Foundation, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. Finally, we thank members from the Bohannan Lab and Green Lab at the University of Oregon for their comments, in particular Brendan Bohannan, Kathryn Docherty, James O’Dwyer, Becky Mueller and Liz Perry.

References


Table 1

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Symbol</th>
<th>Macroecological distribution</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generalized SADs (GSADs)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of individuals measured in a fixed area (density)</td>
<td>$N$</td>
<td>Species individuals distribution (SID)</td>
<td>$S_N$</td>
</tr>
<tr>
<td>Biomass</td>
<td>$M$</td>
<td>Species biomass distribution (SBD)</td>
<td>$S_M$</td>
</tr>
<tr>
<td>Energy use</td>
<td>$E$</td>
<td>Species energy distribution (SED)</td>
<td>$S_E$</td>
</tr>
<tr>
<td><strong>Other macroecological distributions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body-size</td>
<td>$\bar{m}$</td>
<td>Species size distribution (SsD)</td>
<td>$S_{\bar{m}}$</td>
</tr>
<tr>
<td>Individual energy use</td>
<td>$\bar{e}$</td>
<td>Species individual energy distribution (SeD)</td>
<td>$S_{\bar{e}}$</td>
</tr>
</tbody>
</table>
### Table 2

<table>
<thead>
<tr>
<th>General allometry</th>
<th>No error</th>
<th>Log-normally distributed error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_Y(Y) = \frac{S_X\left(f^{-1}(Y)\right)}{</td>
<td>f'(f^{-1}(Y))</td>
<td>} )</td>
</tr>
<tr>
<td>Power-law allometry</td>
<td>( S_Y(Y) = \frac{1}{a} S_X\left(\frac{1}{a}(Y-\log(c))\right) )</td>
<td>( S_Y(Y) = \int \frac{1}{\sigma \sqrt{2\pi}} \exp \left(\frac{(Y-\log(c)-ax)^2}{2\sigma^2}\right) S_X(X) dX )</td>
</tr>
</tbody>
</table>
Figure 1

- birds
- fish
- mammals
- trees
Figure 2

% species

log (N)   log (M)   log (E)

birds
fish
trees
Figure 3

(a)

Species Individuals Distribution (SID)

Species Energy Distribution (SED)

Species Biomass Distribution (SBD)

(b)

Species Individuals Distribution (SID)

Species individual-energy Distribution (ScD)

Species-size Distribution (SsD)
Box 1: Inferring Pattern and Process in Ecological Communities

The theoretical framework outlined in Section 2 can be used as a tool to understand a variety of empirical relationships between macroecological distributions. To demonstrate the utility of this framework, we consider the variance-variance relationship [Eq. 7 and 9]. We highlight three attributes of the variance-variance relationship that allow us to infer pattern and process in ecological communities. Specifically: (i) the position (i.e. coordinates) of a sampled community along the relationship relative to the 1:1 line; (ii) the slope of relationship; and (iii) the intercept of the relationship. For illustrative purposes we examine the $\text{Var}[M] - \text{Var}[N]$ relationship.

To derive the variance-variance relationship, first consider the case where the biomass of a species population is functionally determined by numerical abundance (i.e. a “$N$ causal” model). Using the definition of the average body size within a population $\bar{m} = \frac{M}{N}$, and the expression for the allometric size-density relationship $N \propto \bar{m}^{a_{\text{all}}}$, we obtain: $M \propto N^{\frac{1}{a_{\text{all}}} + \frac{1}{a_{N}}}$. Log-transforming this relationship and allowing for variation in $M$ not explained by $N$ yields:

$$\log M = \log c + \left(1 + \frac{1}{a_{N}} \right) \log N + Z_{M/N}, \quad \text{[Eq. 1]}$$

where $c$ is the proportionality constant in the scaling relationship between $M$ and $N$ and $Z_{M/N}$ captures variation among species biomass that is not explained by population density ($Z_{M/N}$ is a random variable independent of $N$). Taking the variance of both sides reads:
\[ Var(\log M) = \left( 1 + \frac{1}{a_{N|M}} \right)^2 Var(\log N) + Var(Z_{M|N}). \]  \[\text{Eq. 2}\]

Alternatively, assuming a “\(M\) causal” model (where the numerical abundance of a species population is functionally determined by its biomass) yields a different expression for the variance-variance relationship. In this case we model \(N \sim M^{\left(1 + \frac{1}{a_{N|M}}\right)^{-1}}\). Log-transforming this relationship and allowing for variation in \(N\) not explained by \(M\) yields:

\[ \log N = c' + \left( 1 + \frac{1}{a_{N|M}} \right)^{-1} \log M + \log Z_{N|M}, \]  \[\text{Eq. 3}\]

where \(c'\) is the proportionality constant in the scaling relationship between \(N\) and \(M\) and \(Z_{N|M}\) captures variation among species population density that is not explained by population biomass (\(Z_{N|M}\) is a random variable independent of \(M\)). Taking the variance of both sides and rearranging reads:

\[ \text{var}(\log M) = \left( 1 + \frac{1}{a_{N|M}} \right)^2 \text{var}(\log N) - \left( 1 + \frac{1}{a_{N|M}} \right)^2 \text{var}(\log Z_{N|M}) \]  \[\text{Eq. 4}\]

Figure 1 shows the variance-variance relationship described by Equation 2 (or equivalently 4) in the case of perfect allometry (i.e. \(M\) is entirely determined by \(N\) and vice-versa).

The position (or coordinates) of a sampled community on the variance-variance graph reflects ecological pattern. Communities that lie in the shaded area above the 1:1 line
have a less even species-biomass distribution relative to the species-individual distribution (\(N\) is more equitably distributed across species than \(M\)). In contrast, communities that lie in the non-shaded areas below the 1:1 line have a more even species-biomass distribution relative the species-individual distribution (\(M\) is more equitably distributed across species than \(N\)).

The *slope* of the variance-variance relationship also reflects ecological processes as the exponent of the size-density relationship \(a_{\text{Nim}}\) governs its value. For example, the slope is steeper than 1 in communities characterized by \(a_{\text{Nim}} < -\frac{1}{2}\) (which encompasses the predicted “Damuth Rule” \(a_{\text{Nim}} = -\frac{3}{4}\)), but shallower than 1 as soon as \(a_{\text{Nim}} > -\frac{1}{2}\).

The sign of the *intercept* of the variance-variance relationship is dictated by the direction of the functional relationship between \(M\) and \(N\) (Figure 2). The intercept is positive under a “\(N\) causal” model (\(M = f(N)\)) but negative under a “\(M\) causal” model (\(N = f(M)\)). The intercept may thus be used in natural communities to understand causal pathways linking \(M\) and \(N\) and therefore to gain insight into forces organizing communities.
Figure 1: The expected variance-variance relationship between biomass (M) and numerical abundance (N) depends on the exponent of the size-density relationship (here the size-density relationship is assumed exact)
Figure 2: With deviations about the size-density relationship, the expected variance-variance relationship between biomass (M) and numerical abundance (N) depends on the "model of causality"
Box 2: How macroecological distributions may arise from community assembly rules

Macroecological distributions (e.g. GSADs, SsD, SeD) originate from multiple processes acting at different spatial scales. They are often observed at the level of local communities, but local communities are subject to processes that operate at both the local and the regional scale (MacArthur & Wilson, 1967; MacArthur, 1972; Brown & Maurer 1987, Ricklefs & Schluter, 1993). Large-scale determinants (e.g. barriers to dispersal, climate, evolution, extinction) determine the geographic ranges of species, and therefore the potential set of species that can persist in a local community. Based on their traits, some species are locally excluded by competitors or are simply not sufficiently adapted (Tilman 1982, 1990). Local processes therefore restrict the set of possible species. The net effect of local processes in turn restricts the geographic ranges of species: if a species in a region does not possess the right traits to occupy a niche in any local community, it will go regionally extinct (‘upward’ forces of local conditions on the geographic distribution of species). The regularities found in this filtering or ‘sub-setting’ of species groups have been called assembly rules (Eriksson 1993), which set the distribution of the functional traits of species in local communities (McGill et al. 2006). In this mechanistic view, local communities are assembled from the ‘parts on the shelf’ of the geographic region in which they occur by passing through a set of filters. What parts pass through the filters depend on the species’ traits.
The figure below provides an overview of how local macroecological distributions may arise from community assembly rules. The species found at the geographical scale show a particular distribution of trait values, subject to evolutionary processes. Only a subset of those species is expected to reach a particular local community (Loreau et al. 2001). Of these, only a subset may be able to establish. The species that do establish determine the distribution of trait values, such as the species-size distribution. This distribution, together with trait-trait relationships and local environmental conditions (e.g. temperature) in turn determines the individual-energy use distribution. Various biotic interactions among species, and between species and environmental factors, determine the “abundance” of the species present in the local community. This outcome depends critically on the underlying distribution of trait values. Which “abundance” is primarily determined depends on the relative importance of different ecological processes for structuring communities. For example, the SID may be the pattern most directly reflecting processes if birth, death and recruitment are at the core of community organization. Alternatively, the SBD may be the pattern most directly reflecting processes if the community is strongly subject to disturbances affecting species’ biomass. Or the SER might also be the pattern most directly reflecting processes if resource division rules are dominant processes. Because these three GSADs are interrelated, processes acting on one of them will also affect the others.
Figure 1: Hypothesized relation between macroecological distributions and various important processes that structure communities on local and regional (geographic) scales. The different symbols of different size represent individuals of different species, characterized by different traits (e.g. body-size, resource requirements, dispersal ability).
Taking species abundance distributions beyond individuals: Appendix

Appendix S1: Detailed data sources and methods

Birds

We use data from the North American Breeding Bird Survey (BBS; Robbins et al. (16); Sauer et al. (17)), which consists of several thousand survey routes scattered across the continental United States and southern Canada. Data are gathered by volunteer observers who identify and count individuals of every bird species seen or heard at each of 50 stops along a 40 km route. The BBS dataset thus allows for the observation of patterns at both the local scale (individual survey routes) and continental scale (aggregating data across routes). We only use data from the 1400 routes for which surveys were conducted every year over the 5-year period 2002-2006 in order to minimize the chances of failing to detect rare species McGill (12); (Hurlbert and White). We exclude species not well-covered by BBS survey methodology (i.e., waterbirds, raptors, nocturnal species) and focus on 349 species of terrestrial land birds. Since sampling effort is constant across years and routes, we calculate density for each species in each route using the sum of the counts over the 5 year period and at the continental scale using the sum of densities per route. Mean species body-mass measurements are taken from the literature Dunning (5), and per-capita energy use is calculated using empirically derived field metabolic rates (FMR; Nagy et al. (15)) for all bird species: \( \tau \approx FMR \approx 10.5 \, m^{0.681} \).
Fish

Fish data is obtained from a stratified survey of all major drainages in Trinidad, which took place between 1996 and 1998. Data comes from seventy-six sites. The section of stream (average length 50 m) sampled at each site is short enough to be fished thoroughly, yet long enough for all species present to be represented in the catch. The sampling protocol includes major habitat types present in the river at that point (e.g. pool and riffle). Electrofishing is employed where possible, but is replaced by seining (mesh size 1.25 cm) when rivers are turbid. Large deep rivers are sampled with gillnets and a trammel net. Guppies *Poecilia reticulata* and other small fish are collected with dip nets. Sampling effort is consistent across sites. The total number of individuals is recorded for each species at each site. Biomass is measured in the field at the time of fishing for each species and represents the total wet weight of all individuals caught. Per-capita energy use is calculated based on re-fitting data on resting metabolic rates Gillooly et al. (8) using a multiple regression on appropriately transformed data. This relationship accounts for variation in local temperature as well as body size: $e \approx MR \approx 31382 m^{0.7526} e^{-0.4319 kT}$ where k is Boltzman’s constant (8.6 × 10$^{-5}$ eV K$^{-1}$) and T is temperature measured in degrees Kelvin. Temperature is calculated at each site from the average of three temperatures measurements recorded at the beginning, middle, and end of sampling. The value chosen for temperature has in practice no effect on our results since it is the same for every species within local communities.

Mammals

We use data from several small mammal communities from the Sevilleta LTER in New Mexico (6) and the Portal Project in Arizona Brown (1); Ernest et al. (7). These studies include individual measurements of size and thus biomass can be directly calculated by summation without relying on mean species values (contrary to the bird data). The Sevilleta data comes from six sets of mark-recapture webs sampled continuously from 1994 to 1998 (Five Points Grass, Five Points Larrea, Goat Draw, Rio Salado Grass, Rio Salado Larrea and Two 22). Data is summed over three days within each census, the two annual censuses, and over the five year period. Recaptures within a single census is excluded. We use data from the control plots of the Portal Project (see Brown
1998 for details of the study) and sum the values of numerical abundance and biomass over the 12 monthly censuses and the five years from 1994-1998. Per-capita energy use for both studies is calculated from the allometry reported in Nagy et al. (1999) relating field metabolic rate (FMR) and mass for desert mammals: $\bar{m} \approx FMR \approx 3.18 m^{0.785}$.

**Trees**

We use data on trees from the Center for Tropical Forest Studies network (http://www.ctfs.si.edu). Within a 50-ha plot in Barro Colorado Island in central Panama, spatial location, species identification and diameter at breast height ($d$) are reported for every stem $>1cm$ Condit (4); Hubbell et al. (10, 9). To estimate individual aboveground mass we use the empirical interspecific allometry $m = 0.124 d^{2.53}$ relating individual mass ($m$) to diameter at breast height Brown (2); we then sum the individual tree masses to obtain an estimate of biomass. More accurate species-specific allometries incorporating wood density are available for BCI Chave et al. (3), but the allometry we use provides a first good approximation close to the theoretical prediction $m \propto d^{\frac{8}{3}}$ West et al. (18). The energetics of trees *stricto senso* is not well characterized, but it has been proposed that surrogates such as biomass production, water consumption or respiration rates scale as $m^3$. West et al. (18) (thus $d^2$) and foresters have traditionally assumed that total basal area is a decent measure of total production or resource use. To estimate per-capita energy use, we use the relationship derived by refitting the data from Gillooly et al. (2001) using a multiple regression on appropriately transformed data: $e \approx respiration \approx e^{16.949} m^{0.692} e^{-0.642} kT$, where k is Boltzman’s constant ($8.6 \times 10^{-5} eV K^{-1}$) and T is temperature measured in degrees Kelvin. Temperature is calculated as the mean annual temperature (average within days, then within months and then the whole year) averaged over the last 5 years. The value chosen for temperature has in practice no effect on our results since it is the same for every species. The data available for estimating the allometric relationship of energy use by plants is based on seedlings and plant parts Gillooly et al. (8), and is thus necessarily a rough estimation. In addition, photosynthesis is strongly dependent on light availability, and light availability is highly dependent on size in tropical forest: small trees in the shaded understory are likely not at their maximum metabolic rate (13; 14). As this is the best data
available, it is the one we use to get a reasonable though course characterization of the energy use of tree species.

Appendix S2: Detailed empirical results

Evenness

The bird data consist of 1400 local communities. Individuals are distributed with more equitability than biomass in all local communities but 5 and with more equitability than energy use in all local communities but 15. Energy use is more equitably distributed than biomass in all local communities. The fish data consist of 76 local communities. Individuals are distributed with more equitability than biomass in 55 of these 76 local communities and with more equitability than energy use in 43 of them. Energy use is more equitably distributed than biomass in 59 of the 76 local communities. The mammal data consist of 7 local communities. Individuals are distributed with more equitability than biomass in all of them and with more equitability than energy use in all local communities but 1. Energy use is more equitably distributed than biomass in all local communities. The tree data consist of 50 local communities. Individuals are distributed with more equitability than biomass and energy use in all of them. Energy use is more equitably distributed than biomass in all local communities.

Skewness

Local bird communities are slightly more often characterized by a left-skewed SID (814 out of 1400, ~ 58%), but right-skewed SBDs and SEDs (~75% and 59% of the local communities respectively). Local fish communities are characterized by mostly right-skewed SIDs (48 out of 76, ~ 63%) and slightly more right-skewed SEDs (~ 54%), but slightly more left-skewed SBDs (~55%). Five out of the seven local mammal communities show a positive skew for the SID, the SBD and the SED, while the two other local communities show a negative skew for the three distributions. All the local 1 ha local tree plots show a positively-skewed SID, and all of them a negatively-skewed SBD (but 2 plots) and SED.
Appendix S3: Detailed derivations relevant to the conversion framework

General conversion formula

Macroecological distributions are all interrelated by conditional probabilities (or probability densities; e.g. Figure S1). The conversion between the frequency distribution $S_X$ and the frequency distribution $S_Y$ is given by the general formula:

$$ S_Y (Y) = \int P (Y|X) S_X (X) dX $$

(1)

With

$$ Y = f(X) + \varepsilon $$

where $f$ is a general allometry and the error $\varepsilon$ is independent of $X$ and centered ($E(\varepsilon) = 0$), $P (Y|X)$ reads:

$$ P(Y|X) = P(f(X)+\varepsilon|X) = P(\varepsilon|X) = P_\varepsilon(\varepsilon) = P_\varepsilon(Y-f(X)) $$

Combining with 1, the conversion from $X$ to $Y$ is given by the general formula:

$$ S_Y (Y) = \int P_\varepsilon(Y-f(X)) S_X (X) dX $$

(2)

Specific conversion formulas (Table 2)

- General allometry, no error

When there is no error around an allometric relationship $f$, $P_\varepsilon$ may be written as a Dirac delta function:

$$ P_\varepsilon(Y-f(X)) = \delta(Y-f(X)) $$
For any monotonic function $g$ with root $x_i$:

$$\delta (g(x)) = \frac{\delta (x - x_i)}{|g'(x_i)|}$$

where the prime denotes the derivative. We thus have (for $f'(X) \neq 0$):

$$\delta (Y - f(X)) = \frac{\delta (X - f^{-1}(Y))}{|f'(X)|}$$

Therefore, (2) becomes

$$S_Y(Y) = \frac{S_X(f^{-1}(Y))}{|f'(f^{-1}(Y))|}$$

(3)

- **Power-law allometry, no error**

Substituting in 3, for a power-law allometry $f : x \rightarrow \log (c) + ax$ without error:

$$S_Y(Y) = \frac{1}{|a|} S_X \left( \frac{1}{a} (Y - \log (c)) \right)$$

(4)

- **General allometry, normally distributed error**

Substituting in 2, for a normally distributed error $P_\varepsilon(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{x^2}{2\sigma^2} \right)$ and a general allometric relationship $f$:

$$S_Y(Y) = \int \frac{1}{\sigma \sqrt{2\pi}} \exp \left( \frac{(Y - f(X))^2}{2\sigma^2} \right) S_X(X) dX$$

(5)

- **Power-law allometry, normally distributed error**

Substituting in 5, for a normally distributed error and power-law allometry:

$$S_Y(Y) = \int \frac{1}{\sigma \sqrt{2\pi}} \exp \left( \frac{(Y - \log (c) - aX)^2}{2\sigma^2} \right) S_X(X) dX$$

(6)
The geometric SID results from the uniform SsD (case power-law allometry with no error)

It has been shown by Loehle (11) that the geometric SID results from the uniform SsD when an exact power-law allometry is assumed. This result is easily reproducible using our framework. Assume that the size distribution is uniform on a log scale, i.e:

\[ S_s(\log(\bar{m})) = \begin{cases} \frac{1}{\log(\bar{m}_{\text{max}}) - \log(\bar{m}_{\text{min}})} & \text{if } \bar{m}_{\text{min}} \leq \bar{m} \leq \bar{m}_{\text{max}} \\ 0 & \text{otherwise} \end{cases} \]

substituting in 4 leads (with \( a_{N|\bar{m}} \leq 0 \))

\[ S_N(\log(N)) = \begin{cases} \frac{1}{\log(\bar{m}_{\text{max}}) - \log(\bar{m}_{\text{min}})} & \text{if } c\bar{m}_{\text{max}}^{a_{N|\bar{m}}} \leq N \leq c\bar{m}_{\text{min}}^{a_{N|\bar{m}}} \\ 0 & \text{otherwise} \end{cases} \]

i.e. the species abundance distribution is also uniform on a log-scale. Note that the geometric rank abundance curve is equivalent to the uniform distribution on a log scale, and we therefore reproduce the results by Loehle (2006). The advantage of our framework is that it offers the possibility to extend such predictions to biologically more realistic cases, in particular incorporating the effect of variation around the size-density relationship.

Converting the skewness of macroecological distributions

We write

\[ Y = \log(c_{Y|X}) + a_{Y|X}X + Z_{Y|X} \quad (7) \]

with \( Z_{Y|X} \) independent of \( X \). Assume that \( Z_{Y|X} \) is centered (\( E[Z_{Y|X}] = 0 \)). The expected value of \( Y \) reads:

\[ E[Y] = E[\log(c_{Y|X}) + a_{Y|X}X + Z_{Y|X}] = \log(c_{Y|X}) + a_{Y|X}E[X] + E[Z_{Y|X}] = \log(c_{Y|X}) + a_{Y|X}E[X] \]
The variance of $Y$ reads:

$$V[Y] = a^2_{Y|X} V[X] + V[Z_{Y|X}]$$

The skewness of $Y$ is defined by:

$$\gamma_1 = \frac{E[(Y - E[Y])^3]}{(E[(Y - E[Y])^2])^{\frac{3}{2}}} = \frac{E[(Y - E[Y])^3]}{V[Y]^{\frac{3}{2}}}$$

$$E[(Y - E[Y])^3] = E\left(a^3_{Y|X} (X - E[X]) + Z_{Y|X}\right)^3 = E\left(a^3_{Y|X} (X - E[X])^3 + Z_{Y|X}^3 + 3a^2_{Y|X} A(X - E[X])^2 X_{Y|X} + 3a_{Y|X} Z_{Y|X} X_{Y|X}\right)$$

$Z_{Y|X}$ and $X$ are independent, so that:

$$E\left[3a^3_{Y|X} (X - E[X])^2 Z_{Y|X}\right] = 3a^2_{Y|X} E[Z_{Y|X}] E[(X - E[X])^2] = 0$$

$$E\left[3Z_{Y|X}^2 a_{Y|X} (X - E[X])\right] = 3a_{Y|X} E[Z_{Y|X}^2] E[(X - E[X])] = 0$$

Finally

$$E[(Y - E[Y])^3] = E\left(a^3_{Y|X} (X - E[X])^3\right) = a^3_{Y|X} E[(X - E(X))^3] + E[Z_{Y|X}]^3$$

Thus the skewness becomes

$$\gamma_1 = \frac{a^3_{Y|X} E[(X - E[X])^3] + E[Z_{Y|X}]^3}{(a^2_{Y|X} V[X] + V[Z_{Y|X}])^{\frac{3}{2}}}$$

This expression shows that $Y$ can be skewed even if $X$ is not, if $Z_{Y|X}$ is skewed. In the case of unskewed $Z_{Y|X}$:

$$\gamma_1 = \frac{a^3 E[(X - E[X])^3]}{(a^2 V[X] + V[Z_{Y|X}])^{\frac{3}{2}}}$$

This formula shows that if $X$ explains $Y$ with a given unexplained variance, the absolute value
of the skewness of $Y$ is always smaller than the absolute value of the skewness of $X$. This formula also shows that the sign of the skewness of $Y$ is opposite to that of $X$ for $a \leq 0$. In particular, with $a_{M|N} = 1 + \frac{1}{a_{N|m}}$, $N$ and $M$ are expected to have opposite skew for $1 + \frac{1}{a_{N|m}} < 0$, or equivalently $0 > a_{N|m} > -1$. With $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|m}}$, $N$ and $E$ are expected to have opposite skew for $1 + \frac{a_{e|m}}{a_{N|m}} < 0$, or equivalently $0 > a_{N|m} > -a_{e|m}$.

**Appendix S4: Equitability in the distribution of individuals, biomass and energy**

Assuming that the relationships between both $N$ and $\bar{m}$ and $\sigma$ and $\bar{m}$ are power-law, we denote $c_{N|m}$ and $a_{N|m}$ the normalization constant and power-law exponent of the allometry between $\bar{m}$ and $N$. Writing $E = N\sigma$ and $M = N\bar{m}$, it is obvious that the relationship between any 2 of the 5 variables $N$, $\bar{m}$, $\sigma$, $M$ and $E$ is also power-law. Notations for the allometries between any two other variables are denoted accordingly. For example if $N \sim \bar{m}^{a_{N|m}}$, then $\bar{m} \sim N^{\frac{1}{a_{N|m}}}$, so that $M \sim N^{\left(1 + \frac{1}{a_{N|m}}\right)}$ and $a_{M|N} = 1 + \frac{1}{a_{N|m}}$. If furthermore we neglect intraspecific body-size variation (reasonable assumption in the case for determinant growers such as birds and mammals), we can write $\sigma \sim \bar{m}^{a_{e|m}}$, so that $E \sim N^{\left(\frac{1}{a_{N|m}}\right)}a_{e|m} \sim N^{1 + \frac{a_{e|m}}{a_{N|m}}}$ and $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|m}}$. For example with Damuth ($a_{N|m} = -0.75$) and Kleiber’s ($a_{e|m} = 0.75$) coefficients: $a_{M|N} = -0.33$ and $a_{E|N} = 0$ (energy equivalence rule).

Using the formulas presented in the manuscript (section 2), it is straightforward to derive the relationships between the variance in number of individuals, biomass and energy use. This is done in Box 1 for number of individuals and biomass. For simplicity in the figures in Box 1 we have separated the effect of the slope of the size-density relationship (case without error) to the effect of the “causal” relationship. In Figure S2a) we combine the two effects to illustrate the conditions under which biomass is expected to be more equitably distributed than individuals. Figure S2b is the parallel for illustrating the conditions under which energy is expected to be more equitably distributed than individuals.
Captions

• Figure S1: *General link between diversity distributions*

We denote probabilities (or probability densities) associated with species level conversions (blue shade, conversion between species individual, biomass and energy distributions) by upper case symbols ($P$), and probabilities associated with per capita level conversions (yellow shade, conversion between species individual, size and per capita energy distributions) by lower case symbols ($p$). Conversions between population level distributions naturally stem from per capita level probabilities (in particular $p(N|m)$ describing the relationship between density and body-size, and $p(e|m)$ describing the relationship between metabolic rate and body-size). Equations in the figure provide the relationships between probabilities at the population and per capita level.

• Figure S2: *Conceptual figure illustrating the effect of allometric slopes, error and causality on the relationships between: a) the equitability of individuals and biomass division b) the equitability of individuals and resource division*

If the error around allometries is ignored (plain lines), whether biomass (or energy) is more equitably distributed than the number of individuals depends on the slope of the size-density relationship (or the ratio of the slopes of the size-density and size-energy relationships, respectively). With Damuth exponent for the size-density relationship (-0.75), biomass is expected to be more equitably distributed than the number of individuals. With Kleiber’s exponent for the size-energy relationship (0.75), energy is expected to be more equitably distributed than the number of individuals for a size-density relationship steeper than -0.375. Incorporating the effect of scatter around the allometries significantly change the results. The relationships between evenness depend on the direction of causality (short-dashed lines versus dashed-point lines).

References


