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., 1986; Sauer et al., 2008), 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, 2003; Hurlbert and White, 2005, 2007). 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, 1993), and per-capita energy use is calculated using empirically derived field metabolic rates (FMR; Nagy et al., 1999) for all bird species: $\overline{e} \approx FMR \approx 10.5\,\overline{m}^{0.681}$.

Fish

Fish data were obtained from a stratified survey of all major drainages in Trinidad, which took place between 1996 and 1998. Data come 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. It is necessary to use a variety of methods to sample the range of habitats found in Trinidad, and in all cases fishing continued until no further individuals were caught. This type of removal sampling (Southwood and Henderson, 2000) is an effective way of sampling stream habitats, and sampling effort can be considered 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. (2001) using a multiple regression on appropriately transformed data. This relationship accounts for variation in local temperature as well as body size: $\overline{e} \approx MR \approx 31382\, \overline{m}^{0.75262} \exp\left(-\frac{0.4319}{kT}\right)$ 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 no meaningful effect on the local scale results because it is the same for every species within each local community.

Mammals

We use data from several small mammal communities from the Sevilleta LTER in New Mexico (Ernest et al., 2000) and the Portal Project in Arizona (Brown, 1998; Ernest, in press) . These studies include individual measurements of body mass 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 the three days within each census, the two annual censuses, and over the five year period. Recaptures within a single census are 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. As all sites are located in the desert southwest, 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: $\overline{e} \approx FMR \approx 3.18\,\overline{m}^{0.785}$.

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 identity and diameter at breast height (d) are reported for every stem $> 1cm$ (Condit, 1998; Hubbell et al., 1999, 2005). 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, 1997); 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., 2003), but the allometry we use provides a first good approximation close to a theoretical prediction $m \propto d^{\frac{8}{3}}$ (West et al., 1997). The energetics of trees *sensu stricto* is not well characterized, but it has been proposed that surrogates such as biomass production, water consumption or respiration rates scale as $m^{\frac{3}{4}}$ West et al. (1997) (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: $e \approx respiration \approx \exp(16.949)\ m^{0.692}\ \exp\left(-\frac{0.642}{kT}\right)$, where k is Boltzman's constant ($8.6 * 10^{-5}\ eV\ K^{-1}$) 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. (2001), 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 (Muller-landau 2006a,b). As this is the best data available, we use it to get a reasonable though coarse characterization of the energy use of tree species.

Appendix S2: Detailed empirical results

We focus here in testing the equivalence of GSADs when communities are considered as a whole (see Zar, 1999; Sokal and Rohlf, 2000; Cojbasic and Tomovic, 2007 for parametric and non-parametric approaches to comparing individual-level communities).

Variance

The bird data consist of 1400 local communities. Individuals are distributed with more equitability than biomass in all local communities but 5 (99.6% of the communities, $p < 0.05$ -run test-), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(M)]}{Var[\log(N)]}$) of $[1.64; 1.67]$. Individuals are distributed with more equitability than energy use in all local communities but 15 (98.9%, $p < 0.05$), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(E)]}{Var[\log(N)]}$) of $[1.32; 1.34]$. Energy use is more equitably distributed than biomass in all local communities ($p < 0.05$), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(M)]}{Var[\log(E)]}$) of $[1.23; 1.24]$. The fish data consist of 76 local communities. Individuals are distributed with more equitability than biomass in 55 of these 76 local communities (72.3%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.9; 5.0]$), and with more equitability than energy use in 43 of them (56.6%, n.s; 95% confidence interval for the factor of deviation: $[0.90; 1.01]$). Energy use is more equitably distributed than biomass in 59 of the 76 local communities (77.6%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[0.95; 6.25]$). The mammal data consist of 7 local communities. Individuals are distributed with more equitability than biomass in all of them (100%, $p < 0.05$; 95% confidence

interval for the factor of deviation: $[1.10; 1.44]$) and with more equitability than energy use in all local communities but 1 (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.03; 1.26]$). Energy use is more equitably distributed than biomass in all local communities (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.06; 1.15]$). The tree data consist of 50 local communities. Individuals are distributed with more equitability than biomass and energy use in all of them (100%, $p < 0.05$; 95% confidence intervals for the factors of deviation: $[3.77; 4.03]$ and $[2.10; 2.23]$ respectively). Energy use is more equitably distributed than biomass in all local communities (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.78; 1.81]$). Taken together, these results demonstrate the non-equivalence of the variance of GSADs when communities are considered as a whole.

Skewness

Local bird communities are more often characterized by a left-skewed SID (814 out of 1400, 58%, $p < 0.05$), but right-skewed SBDs and SEDs (75% and 59% of the local communities respectively, $p < 0.05$). Local fish communities are characterized by mostly right-skewed SIDs (48 out of 76, 63%, $p < 0.05$) and slightly more right-skewed SEDs (54%, n.s.), but slightly more left-skewed SBDs (55%, n.s.). 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 (n.s.). All the local 1 ha local tree plots show a positively-skewed SID, and all of them a negatively-skewed SBD (in the exception of 2 plots) and SED ($p < 0.05$).

Appendix S3: Detailed derivations

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:

$$S_Y(Y) = \int P(Y|X)\, S_X(X)\, dX \tag{1}$$

With

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

where f is a general allometry and the error ε 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:

$$S_Y(Y) = \int P_\varepsilon(Y - f(X))\, S_X(X)\, dX \tag{2}$$

Specific conversion formulas (Table 2)

- *General allometry, no error*

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

$$P_\varepsilon(Y - f(X)) = \delta(Y - f(X))$$

For any monotic 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 $f'(X) \neq 0$):

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

Therefore, (2) becomes

$$S_Y\left(Y\right) = \frac{S_X\left(f^{-1}\left(Y\right)\right)}{\left|f'\left(f^{-1}\left(Y\right)\right)\right|} \tag{3}$$

- *Power-law allometry, no error*

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

$$S_Y\left(Y\right) = \frac{1}{|a|} S_X\left(\frac{1}{a}\left(Y - \log\left(c\right)\right)\right) \tag{4}$$

- *General allometry, normally distributed error*

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

$$S_Y\left(Y\right) = \int \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{\left(Y - f(X)\right)^2}{2\sigma^2}} S_X\left(X\right) dX \tag{5}$$

- *Power-law allometry, normally distributed error*

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

$$S_Y\left(Y\right) = \int \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{\left(Y - \log(c) - aX\right)^2}{2\sigma^2}} S_X\left(X\right) dX \tag{6}$$

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

It has been show Loehle (2006) that the geometric SID results from the uniform SsD when an exact power-law allometry is assumed. This result is easily reproducibe using our framework. Assume

that the size distribution is uniform on a log scale, i.e:

$$S_s\left(\log\left(\overline{m}\right)\right) = \frac{1}{\log(\overline{m}_{max})-\log(\overline{m}_{min})} \qquad if \quad \overline{m}_{min} \leq \overline{m} \leq \overline{m}_{max}$$
$$= 0 \qquad otherwise$$

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

$$S_N\left(\log\left(N\right)\right) = \frac{1}{|a|(\log(\overline{m}_{max})-\log(\overline{m}_{min}))} \qquad if \quad c\overline{m}_{max}^{a_{N|\overline{m}}} \leq N \leq c\overline{m}_{min}^{a_{N|\overline{m}}}$$
$$= 0 \qquad otherwise$$

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.

Variance-variance relationships depend on the choice of "causal model"

There are two ways to describe the functional relationships between X and Y: 1) Y is functionally determined by 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. 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. growth *versus* reproduction response in indeterminate growers such as trees). 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),

$$Y = \log\left(c_{Y|X}\right) + a_{Y|X}X + Z_{Y|X} \tag{7}$$

with $Z_{Y|X}$ independent of X. In this case,

$$Var\left[Y\right] = a_{Y|X}^2 Var\left[X\right] + Var\left[Z_{Y|X}\right] \tag{8}$$

In the second case (X functionally determined by Y)

$$X = \log\left(c_{X|Y}\right) + \frac{1}{a_{Y|X}} Y + Z_{X|Y} \tag{9}$$

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

$$Var\left[Y\right] = a_{Y|X}^2 Var\left[X\right] - a_{Y|X}^2 Var\left[Z_{X|Y}\right] \tag{10}$$

Equations 8 and 10 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. Equations 8 and 10 may be used to understand most relevant "causal models" in natural communities. Indeed, variance values and allometric exponents can be measured from field data, which can then be used to infer the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$. Given that $Var\left[Z_{Y|X}\right]$ is positive, Equation 8 is satisfied if the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$ is positive, indicating a "X causal" model. If instead the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$ is negative, Equation 10 is satisfied, indicating a "Y causal" model.

Converting the skewness of macroecological distributions

We write

$$Y = \log\left(c_{Y|X}\right) + a_{Y|X} X + Z_{Y|X} \tag{11}$$

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

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

The variance of Y reads:

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

The skewness of Y is defined by:

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

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

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

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

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

Finally

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

Thus the skewness becomes

$$\gamma_1 = \frac{a_{Y|X}^3 E[(X - E[X])^3] + E\left[Z_{Y|X}\right]^3}{(a_{Y|X}^2 Var[X] + Var\left[Z_{Y|X}\right])^{\frac{3}{2}}}$$

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

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

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 $a_{M|N} = 1 + \frac{1}{a_{N|\overline{m}}}$, N and M are expected to have opposite skew for $1 + \frac{1}{a_{N|\overline{m}}} < 0$, or equivalently $0 > a_{N|\overline{m}} > -1$. With $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|\overline{m}}}$, N and E are expected to have opposite skew for $1 + \frac{a_{e|m}}{a_{N|\overline{m}}} < 0$, or equivalently $0 > a_{N|\overline{m}} > -a_{e|m}$. With typical values reported for the exponent of the size-density relationship, the direction of the skewness is expected to switch between abundance measures. Under the assumption of unskewed 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, as 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 relationship is shallower than the slope of the size-energy relationship, which is the case in our data and is also supported by others (Brown and Maurer, 1986).

Specific variance-variance relationships (Table 3)

We derive the variance-variance relationship between numerical abundance (N) and biomass (M) under the four simplified scenarios presented in section *Illustration and Empirical Evaluation* (Table 3). The biomass of a species is directly given by $M = N\overline{m}$, where $\overline{m}$ is the average body-size of the species' population. We thus have: $\log(M) = \log(N) + \log(\overline{m})$.

First scenario: M is the explanatory variable; we write $\log(N) = \log(M) - \log(\overline{m})$.

In the first case, the body-size of species has no effect on their ability to accumulate biomass. In other words, M is independent of $\overline{m}$. We find:

$$Var[\log N] = Var[logM] + Var[log\overline{m}]$$

In the second case, we assume a power-law dependency of M on $\overline{m}$ (exponent $a_{M|\overline{m}}$), so that $\log(N) = (1 - \frac{1}{a_{M|\overline{m}}})\log(M) + Z_{N|M}$ with $Z_{N|M}$ independent of M. We find:

$$Var\left[\log N\right] = \left(1 - \frac{1}{a_{M|\overline{m}}}\right)^2 Var\left[logM\right] + Var\left[Z_{N|M}\right]$$

Second scenario: N is the explanatory variable; we write $\log(M) = \log(N) + \log(\overline{m})$.

In the first case, the body-size of species has no effect on numerical abundance. In other words, N is independent of $\overline{m}$. We find:

$$Var\left[\log M\right] = Var\left[logN\right] + Var\left[log\overline{m}\right]$$

In the second case, we assume a power-law dependency of N on $\overline{m}$ (exponent $a_{N|\overline{m}}$), so that $\log(M) = (1 + \frac{1}{a_{N|\overline{m}}})\log(N) + Z_{M|N}$ with $Z_{M|N}$ independent of N. We find:

$$Var\left[\log M\right] = \left(1 + \frac{1}{a_{N|\overline{m}}}\right)^2 Var\left[logN\right] + Var\left[Z_{M|N}\right]$$

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

Assuming that the relationships between both N and $\overline{m}$ and $\overline{e}$ and $\overline{m}$ are power-law, we denote $c_{N|\overline{m}}$ and $a_{N|\overline{m}}$ the normalization constant and power-law exponent of the allometry between $\overline{m}$ and N. Writing $E = N\overline{e}$ and $M = N\overline{m}$, it is obvious that the relationship between any 2 of the 5 variables N, $\overline{m}$, $\overline{e}$, M and E is also power-law. Notations for the allometries between any two other variables are denoted accordingly. For example if $N \sim \overline{m}^{a_{N|\overline{m}}}$, then $\overline{m} \sim N^{\frac{1}{a_{N|\overline{m}}}}$, so that $M \sim N^{\left(1+\frac{1}{a_{N|\overline{m}}}\right)}$ and $a_{M|N} = 1 + \frac{1}{a_{N|\overline{m}}}$. If furthermore we neglect intraspecific body-size variation (reasonable assumption in the case for determinant growers such as birds and mammals), we can write $\overline{e} \sim \overline{m}^{a_{e|m}}$, so that $E \sim N\left(N^{\frac{1}{a_{N|\overline{m}}}}\right)^{a_{e|m}} \sim N^{1+\frac{a_{e|m}}{a_{N|\overline{m}}}}$ and $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|\overline{m}}}$. For example with Damuth ($a_{N|\overline{m}} = -0.75$) and Kleiber's ($a_{e|m} = 0.75$) coefficients: $a_{M|N} = -0.33$ and $a_{E|N} = 0$ (the 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. These formulas depend on assumptions regarding which variable is explanatory vs. explained. Figure S3a illustrates the conditions under which biomass is expected to be more equitably distributed than individuals, and Figure S3b 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 $p\left(N|\overline{m}\right)$ describing the relationship between density and body-size, and $p(\overline{e}|\overline{m})$ 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: *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 within a species are assumed of the same size, so that $m = \overline{m}$ and $e = \overline{e}$. 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 $\log(N) - \log(\overline{m})$ plane (panel *b*) is the size-density relationship, and the projection on the $\log(\overline{e}) - \log(\overline{m})$ 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 energetic 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*).

- Figure S3: *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

Brown, J. H. (1998). *The desert granivory experiments at Portal*. Oxford University Press, New York.

Brown, J. H. and Maurer, B. A. (1986). Body size, ecological dominance, and cope's rule. *Nature*, 324:248–250.

Brown, S. (1997). *Estimating biomass and biomass change of tropical forests.* Food and Agricultural Organization of the United Nations Forestry Paper 134.

Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., and Hubbell, S. P. (2003). Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J. Ecol.*, 91:240–252.

Cojbasic, V. and Tomovic, A. (2007). Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples. *Computational Statistics and Data Analysis*, 51(12):5562–5578.

Condit, R. (1998). *Tropical Forest Census Plots.* Springer-Verlag and R. G. Landes Company, Berlin and Georgetown.

Dunning, J. (1993). *Handbook of Avian Body Masses.* CRC Press, Boca Raton.

Ernest, S. K. M., Brown, J. H., and Parmenter, R. R. (2000). Rodents, plants, and precipitation: Spatial and temporal dynamics of consumers and resources. *Oikos*, 88:470–482.

Ernest, S. K. M., Valone, T. J., and Brown, J. H. (in press). Long-term monitoring and experimental manipulation of a Chihuahuan desert ecosystem near Portal AZ. *Ecology.*

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., and Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science*, 293.

Hubbell, S., Condit, R., and Foster, R. (2005). Barro Colorado Forest Census Plot Data.

Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., Wright, S. J., and Loo de Lao, S. (1999). Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, 283:554–557.

Hurlbert, A. H. and White, E. P. (2005). Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecol. Lett.*, 8(3):319–327.

Hurlbert, A. H. and White, E. P. (2007). Ecological correlates of range occupancy in North American breeding birds. *Global Ecol. Biogeogr.*, 16:764–773.

Loehle, C. (2006). Species abundance distributions result from body size-energetics relationships. *Ecology*, 87(9):2221–2226.

McGill, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature*, 422:881–885.

Muller-landau, H. (2006a). Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecol. Lett.*, 9:575–588.

Muller-landau, H. (2006b). Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol. Lett.*, 9:589–602.

Nagy, K. A., Girard, I. A., and Brown, T. K. (1999). Energetics of free-ranging mammals, reptiles, and birds. *Annu. Rev. Nutr.*, 19(1):p247 – 277.

Robbins, C. S., Bystrak, D., and Geissler, P. H. (1986). *The breeding bird survey: its first fifteen years, 1965-1979*. Resource Publication 157. U.S. Department of the Interior, Fish and Wildlife Service.

Sauer, J. R., Hines, J. E., and Fallon, J. (2008). *The North American Breeding Bird Survey, Results and Analysis 1966 - 2007*. Version 5.15.2008. USGS Patuxent Wildlife Research Center.

Sokal, R. and Rohlf, F. (2000). *Biometry: the principles and practice of statistics in biological research*. Freeman New York.

Southwood, R. and Henderson, P. (2000). *Ecological Methods*. Blackwell Publishing.

Taper, M. L. and Marquet, P. A. (1996). How do species really divide resources? *Am. Nat.*, 147(6):1072–1086.

West, G. B., Brown, J. H., and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276:122–126.

Zar, J. (1999). *Biostatistical analysis*. Prentice Hall Upper Saddle River, NJ.

Figure S1

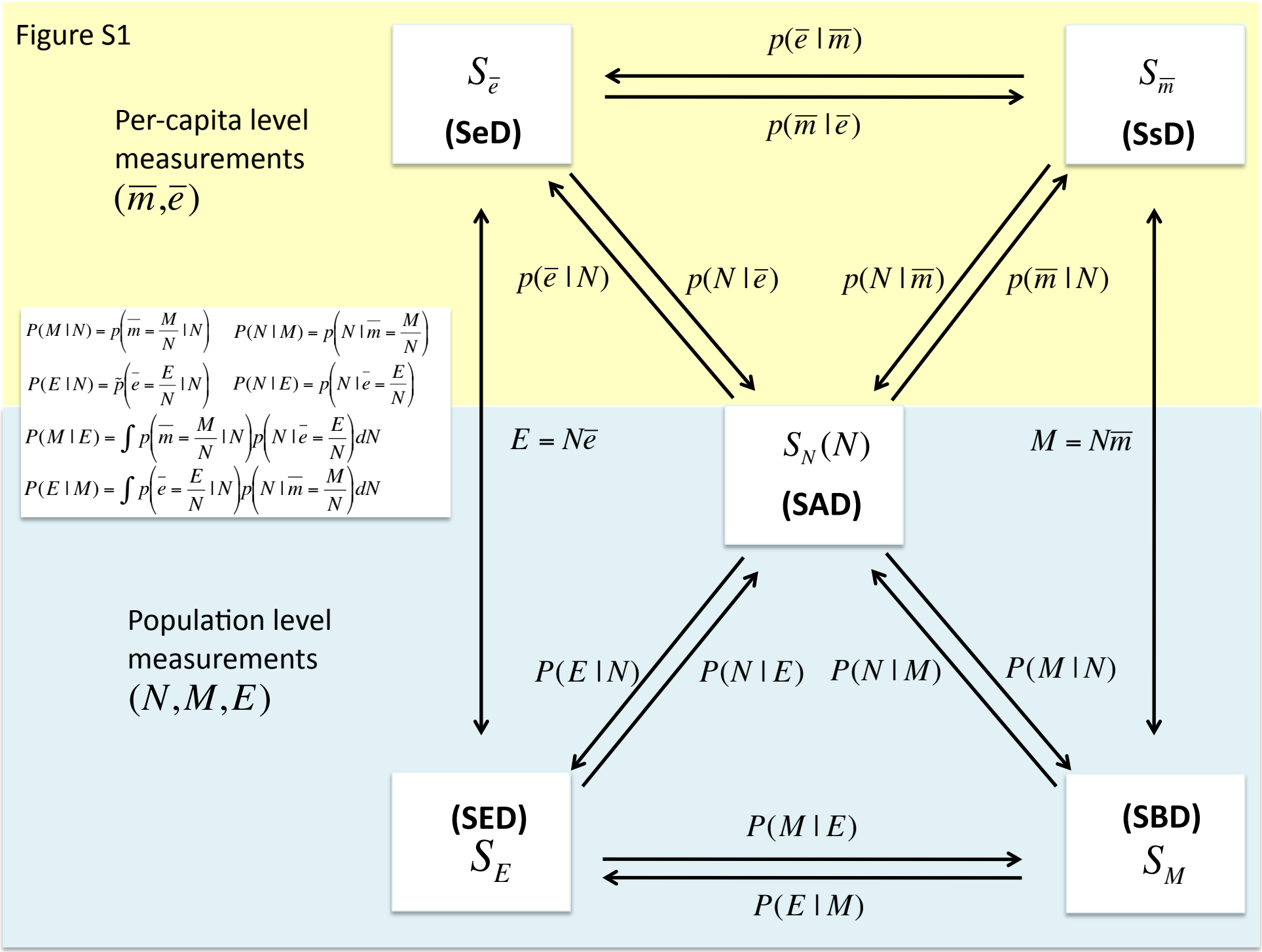
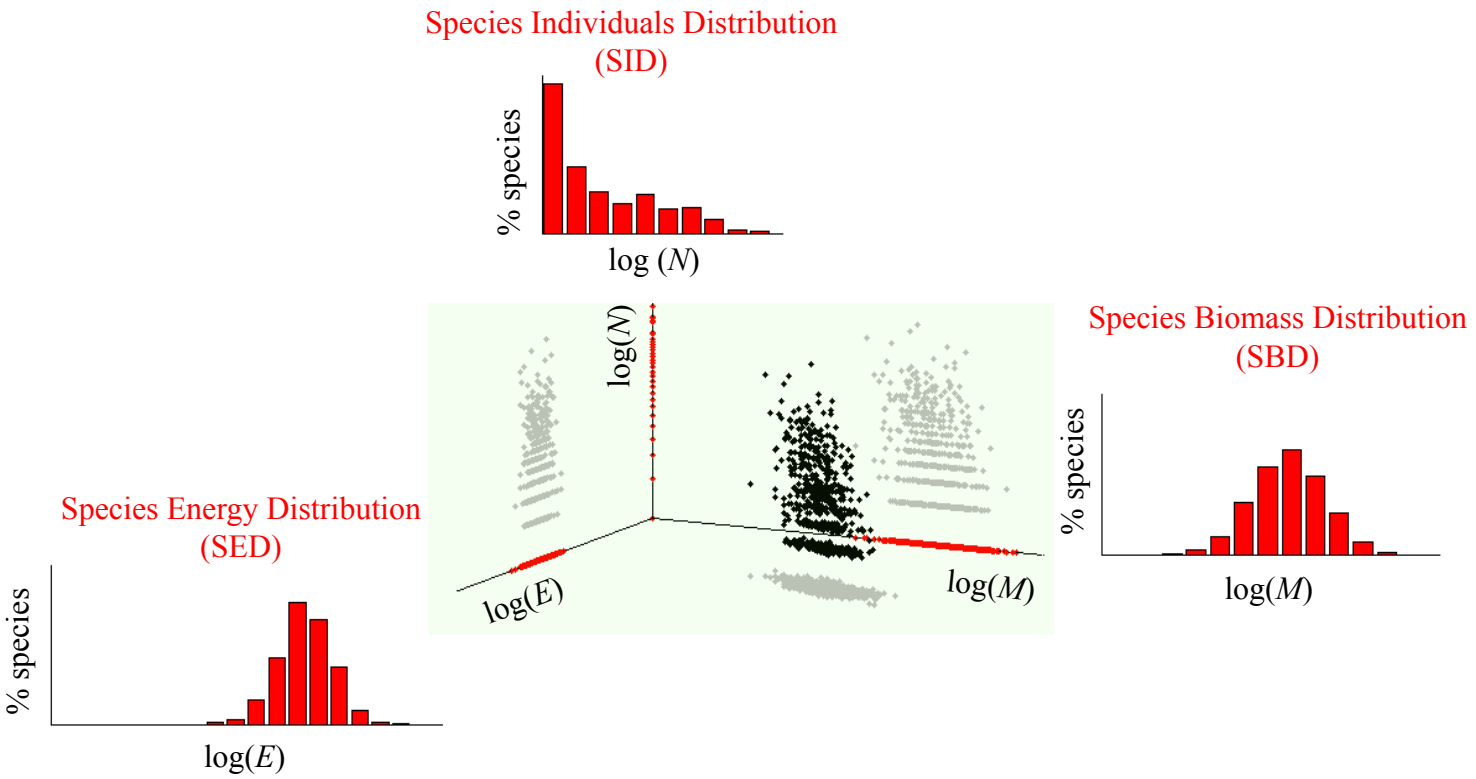

Per-capita level measurements $(\overline{m},\overline{e})$
$S_{\overline{e}}$ (SeD)
$S_{\overline{m}}$ (SsD)
$p(\overline{e}\mid\overline{m})$
$p(\overline{m}\mid\overline{e})$
$p(\overline{e}\mid N)$
$p(N\mid\overline{e})$
$p(N\mid\overline{m})$
$p(\overline{m}\mid N)$
$P(M\mid N)=p\left(\overline{m}=\frac{M}{N}\mid N\right)$ $P(N\mid M)=p\left(N\mid\overline{m}=\frac{M}{N}\right)$
$P(E\mid N)=\tilde{p}\left(\overline{e}=\frac{E}{N}\mid N\right)$ $P(N\mid E)=p\left(N\mid\overline{e}=\frac{E}{N}\right)$
$P(M\mid E)=\int p\left(\overline{m}=\frac{M}{N}\mid N\right)p\left(N\mid\overline{e}=\frac{E}{N}\right)dN$
$P(E\mid M)=\int p\left(\overline{e}=\frac{E}{N}\mid N\right)p\left(N\mid\overline{m}=\frac{M}{N}\right)dN$
$E=N\overline{e}$
$S_N(N)$ (SAD)
$M=N\overline{m}$
Population level measurements (N,M,E)
$P(E\mid N)$
$P(N\mid E)$
$P(N\mid M)$
$P(M\mid N)$
(SED) S_E
(SBD) S_M
$P(M\mid E)$
$P(E\mid M)$

Figure S2

(a)


Species Individuals Distribution
(SID)
% species
log (N)
log(N)
log(E)
log(M)
Species Biomass Distribution
(SBD)
% species
log(M)
Species Energy Distribution
(SED)
% species
log(E)

(b)

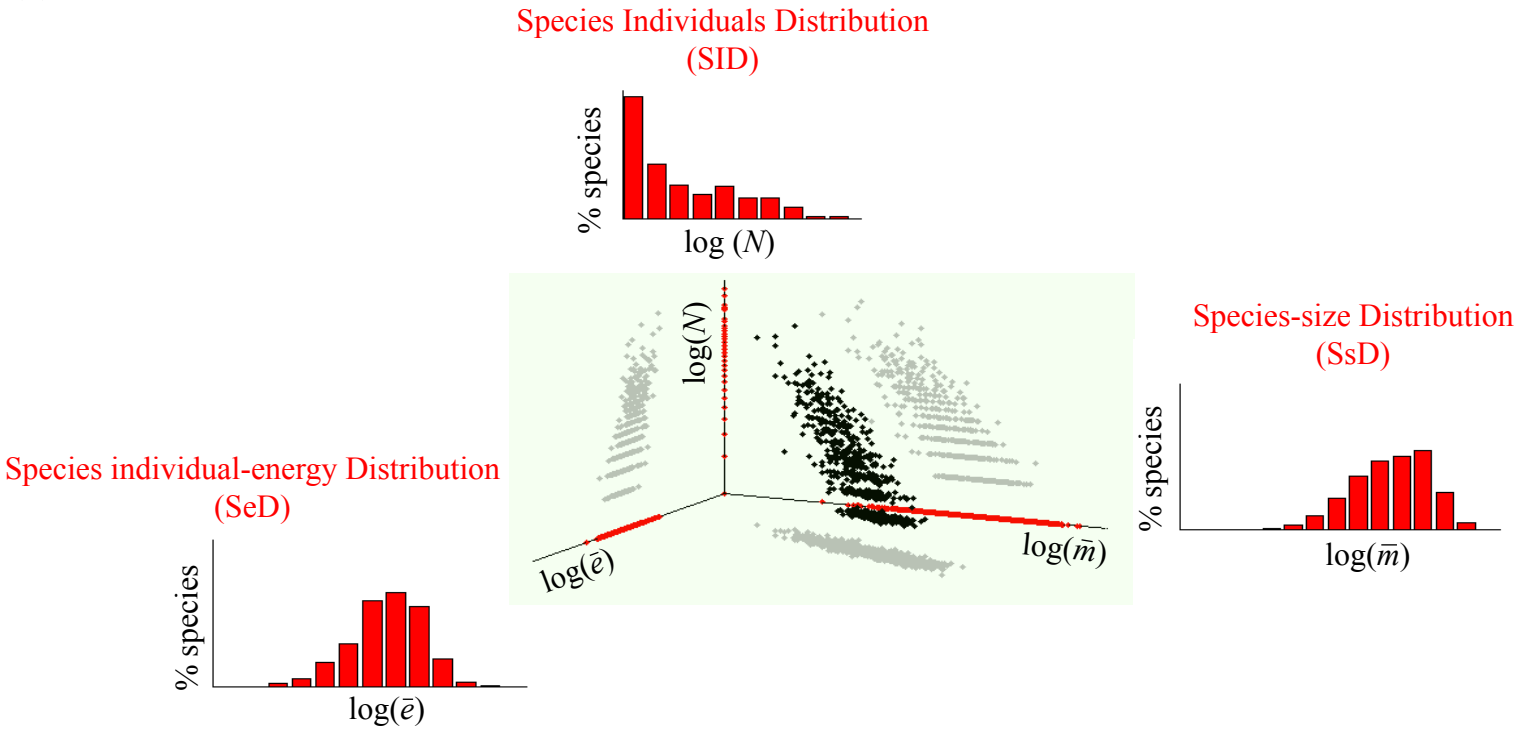

Species Individuals Distribution
(SID)
% species
log (N)
log(N)
log(ē)
log(m̄)
Species-size Distribution
(SsD)
% species
log(m̄)
Species individual-energy Distribution
(SeD)
% species
log(ē)

Figure S3a

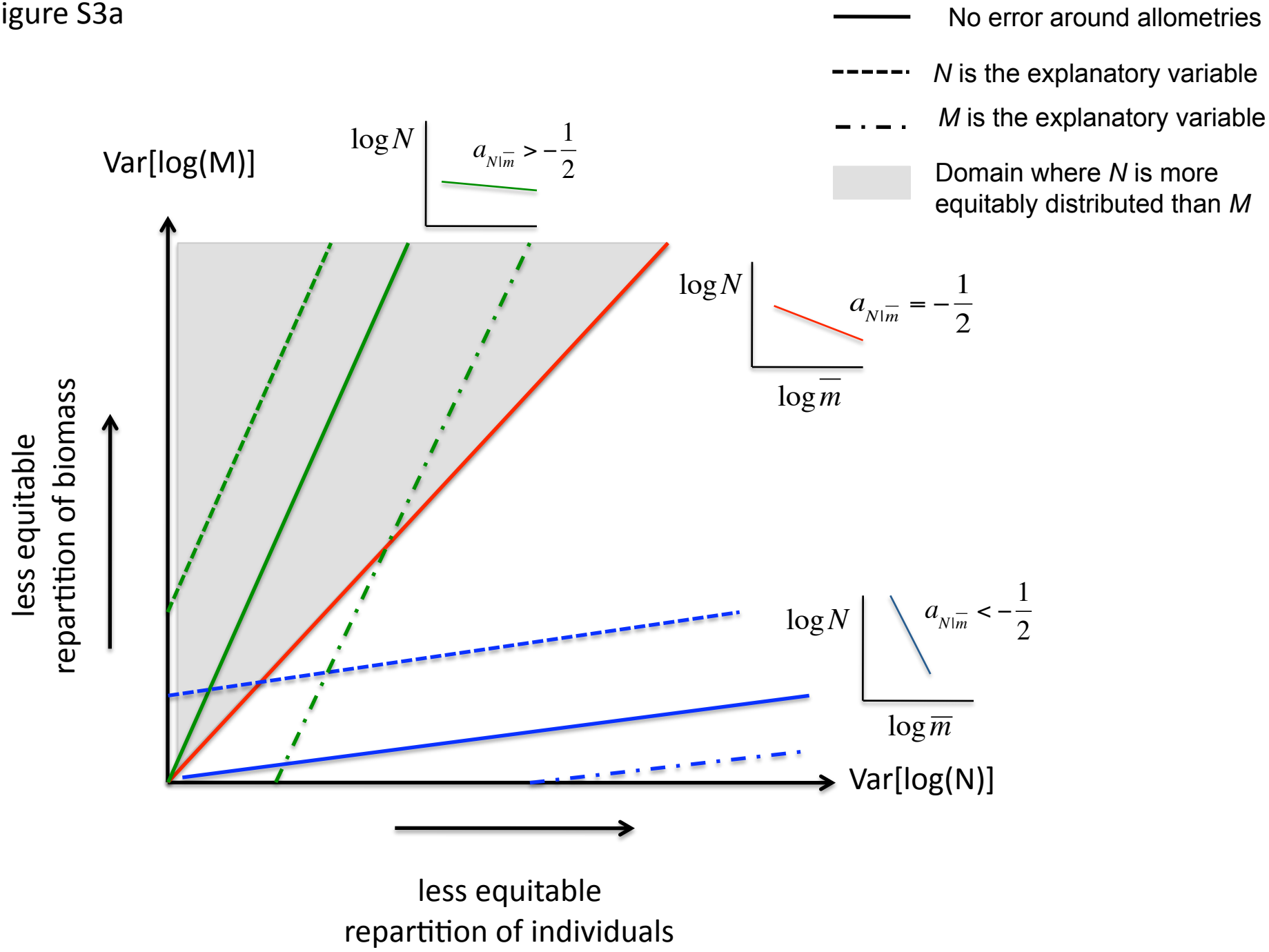
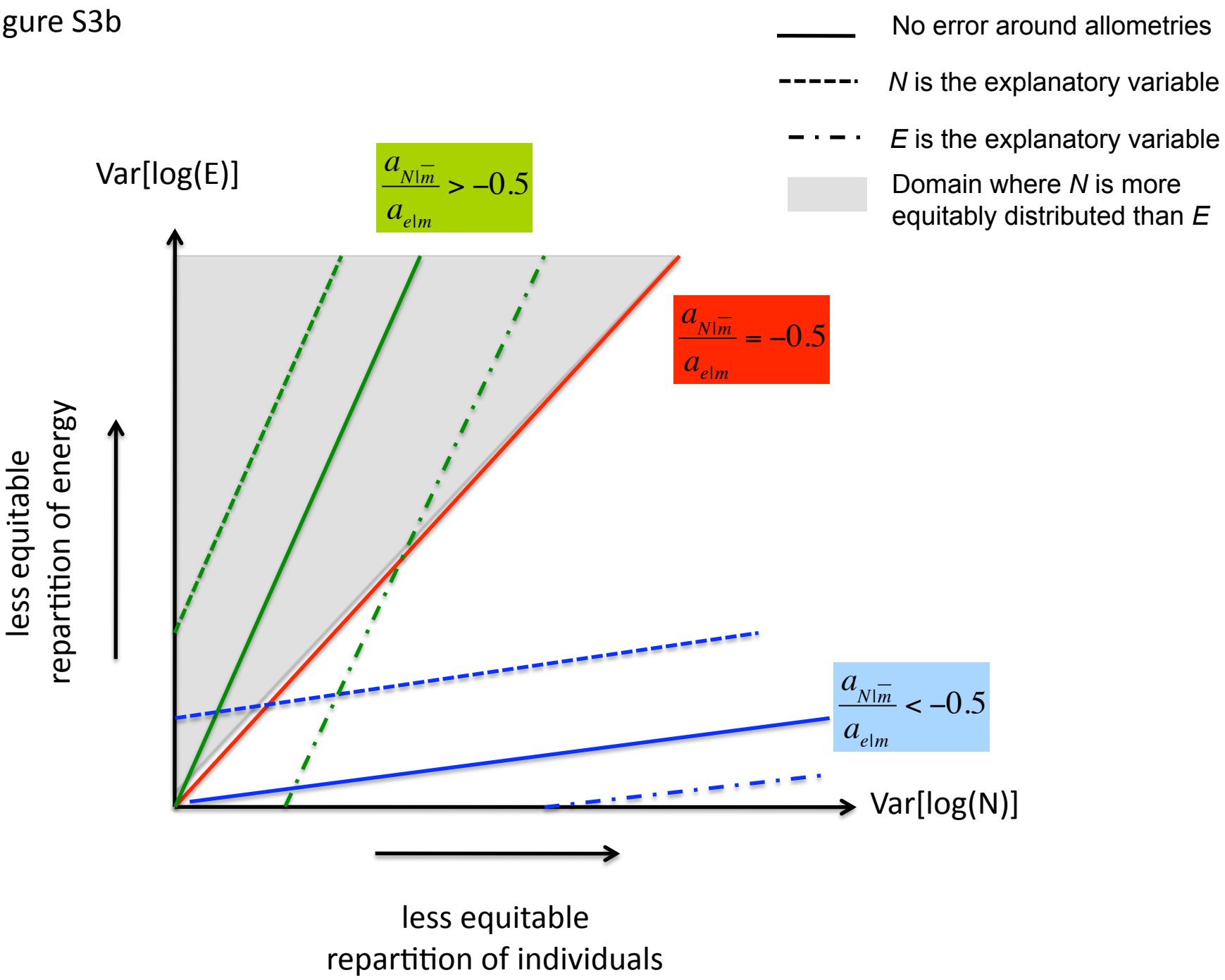

No error around allometries
N is the explanatory variable
M is the explanatory variable
Domain where N is more equitably distributed than M
Var[log(M)]
log N
$a_{N|\overline{m}} > -\frac{1}{2}$
log N
$a_{N|\overline{m}} = -\frac{1}{2}$
log $\overline{m}$
log N
$a_{N|\overline{m}} < -\frac{1}{2}$
log $\overline{m}$
less equitable repartition of biomass
Var[log(N)]
less equitable repartition of individuals

Figure S3b


No error around allometries
N is the explanatory variable
E is the explanatory variable
Domain where N is more equitably distributed than E
Var[log(E)]
$\frac{a_{N|\bar{m}}}{a_{e|m}} > -0.5$
$\frac{a_{N|\bar{m}}}{a_{e|m}} = -0.5$
$\frac{a_{N|\bar{m}}}{a_{e|m}} < -0.5$
less equitable repartition of energy
Var[log(N)]
less equitable repartition of individuals