This is a repository copy of *How to allow SAR collapse across local and continental scales: a resolution of the controversy between Storch et al. (2012) and Lazarina et al. (2013)*.

Article:

How to allow SAR collapse across local and continental scales: a resolution of the controversy between Storch et al. (2012) and Lazarina et al. (2013)

Arnošt Leoš Šizling[1], Eva Šizlingová[1], Even Tjørve[2], Kathleen M.C. Tjørve[2], William E. Kunin[3]

[1]Center for Theoretical Study, Jilská 1, 110 00, Praha 1, Czech Republic; sizling@cts.cuni.cz; eva.sizlingova@seznam.cz

[2]Lillehammer University College, P.O.Box 952, NO-2604 Lillehammer, Norway; even.tjorve@hil.no; kathy.tjorve@hil.no.

[3]School of Biology, LC Miall Building, University of Leeds, Leeds LS2 9JT, UK, W.E.Kunin@leeds.ac.uk

Abstract

Up-scaling species richness from local to continental scales is an unsolved problem of macroecology. Macroecologists hope that proper up-scaling can uncover the hidden rules that underlie spatial patterns in species richness, but a machinery to up-scale species richness also has a purely practical side at the scales and for the habitats where direct observations cannot be performed. The species—area relationship (SAR) could provide a tool for up-scaling, but no valid method has yet been put forward. Such a method would have resulted from Storch et al.'s (2012) suggestion that there is a universal curve to which each rescaled SAR collapses, if Lazarina et al. (2013) had not shown that it does not: both arguments were supported by data analyses. Here we present an analytical model for mainland SAR and argue in favour of the latter authors. We identify (i) the variation in mean species-range size, (ii) the variation in forces that drive SAR at various scales, and (iii) the finite-area effect, as the reasons for the absence of collapse. Finally, we suggest a rescaling that might fix the problem. We conclude, however, that ecologists are still far from finding a practical, robust and easy-to-use solution for up-scaling species richness from SARs.

Introduction

The behaviour of the relationship between species richness and area of the sample plot, the species—area relationship (SAR), is a long standing puzzle in biogeography and ecology. Many fields of biology would profit from a method that could reliably predict the SAR for an unknown region. For example, (i) paleoecology would profit from the knowledge of how to up-scale species richness from findings of small-scaled excavations to total species numbers of the region; (ii) monitoring programmes would be able to assess multi-scale biodiversity patterns and dynamics from local community sampling and (iii) conservation planners could design the optimal size and number of sanctuaries using the relationship between their size and species richness (Simberloff and Abele 1976, Tjørve 2010).

There have been many attempts to find a spatially and taxonomically general curve (e.g., Preston 1962, Rosindell and Cornell 2007, Harte et al. 2009), which can predict any SAR from only a few parameters for the region. Storch et al. (2012) suggested that such a universal SAR exists and that each observed SAR will emerge from the universal SAR by rescaling both axes. Specifically, this entailed multiplying the x-axis by the mean size of the species' geographic ranges, and then multiplying the y-axis by the expected number of species in a sample plot that is as large as the mean species' range. In other words, Storch et al. (2012) proposed that the rescaling

$$S^*(\sigma^*) = \frac{S(\bar{a} \cdot \sigma^*)}{S(\bar{a})} \tag{1}$$

(where $S(\sigma)$ is the observed SAR and $\sigma^* = \sigma/\bar{a}$, i.e, $S(\sigma) = S(\bar{a} \cdot \sigma^*)$; $\bar{a}$ is mean range size; σ and σ^* are the area of sample plot and the rescaled area of sample plot, respectively; and S^* is the rescaled species richness) makes each SAR collapse to a universal SAR that is taxon and continent invariant. According to Storch et al. (2012), this SAR property results from a roughly random placement of species' ranges.

Lazarina et al. (2013) tested the theory of a collapse onto a universal SAR using plant, butterfly and bird data but reported a lack of predicted collapse for SARs from small geographic areas. They did not expect this result and argued that a collapse should be equally expected for data from small areas and whole continents. The reason was that 'there is a large body of theoretical studies arguing how the properties of SAR can be generated by the same processes across different scales (Allen and White 2003, Rosindell and Cornell 2007, Harte et al. 2009)'. The term 'process' or 'underlying mechanism' is, however, used ambiguously in macroecological literature. It may either mean the biological processes (or forces) that influence the location of individuals within the areas where they are observed, or the geometric processes that translate the location of individuals into the observed patterns. Both of these processes must be taken into account when rescaling.

Ten years before the paper by Lazarina et al. (2013), Allen and White (2003) introduced a computer-based simulation of a map of species ranges. Ranges, sample plots, and extent of sampling were all modelled as circles. In double logarithmic space, the SARs computed using the simulated maps were downward accelerating for small sample plots, upward accelerating for large sample plots, and nearly linear for the sample plots in between, thus supporting the hypothesis of a three-phased SAR (Allen and White 2003). The processes modelled by this approach were geometric ones in the sense of the abovementioned discussion, and the model illustrates how different forces affect the SAR at different scales. Intuitively, the scale dependence of the geometric drivers of the SAR suggests a need for a different rescaling at each scale if the aim is a collapse of the SARs onto a universal curve. Nevertheless, this does not mean that the reasoning by Lazarina et al. (2013) and the expectation proposed by Storch et al. (2012) are necessarily wrong in terms of the biological processes. These biological processes may act in the same way at all geographical extents, but they actually combine with geometric processes that vary between scales. This prevents the SAR from collapsing if a simple, scale invariant, rescaling is used.

Here we extended the model by Allen and White (2003). Our model will show that the rescaling by Storch et al. (2012) does not produce a collapse: neither for small nor large scales. We will demonstrate analytically how geometric processes produce variation in the slope of the rescaled SARs, and that this variation prevents SARs from collapsing. The evidence that rescaling, as it is suggested by Storch et al. (2012), does not cause a universal collapse onto a single curve (which Lazarina et al. 2013 shows empirically) undermines the potential of this method for up-scaling species richness. We therefore suggest an alternative rescaling approach that eliminates the geometric effects. This rescaling results in the collapse of 24 out of 26 SARs presented in Storch et al. (2012) and Lazarina et al. (2013). Our results suggest that biological processes may be independent from the extent of the area under examination, although the actual observed patterns vary between local and continental scales.

Theory

In this section, we introduce an analytical function of the SAR based on geometric processes. The mathematical form of the function identifies the geometric constraints, while its parameters capture the biological processes. This function can be applied to separate the biological and geometric processes from each other. We then use this function to rescale the SARs used by Storch et al. (2012) and Lazarina et al. (2013), in order to show how geometric processes prevented these SARs from collapsing. The fact that the rescaled SARs collapsed when we removed the geometric effects from them confirms the macroecological intuition behind a collapse as presented in Storch et al. (2012) and Lazarina et al. (2013).

We first show how a SAR can be expressed in terms of a ratio between two areas, where the area in the denominator (below the fraction line) is given only by the shape and size of the sample plot (hereafter σ) and the geographical extent under the survey (hereafter the arena). The area in the numerator (the expression above the fraction line) is then a sum of areas across species spatial ranges and therefore only the coefficients in the numerator are affected by biological processes. Most of the Theory section discusses how σ affects the areas in the numerator and the denominator. The central point of our computation is that any change in σ affects the scale of resolution.

Coleman (1981) notes that the SAR may be generated as the sum of many species' occupancy—area relationships (OARs): the proportion of samples occupied by a species as a function of σ (see also Williams 1995; Muriel and Mangel 1999; He and Legendre 2002). Although this result was attributed to the random placement of individuals within the arena by Coleman (1981) and his followers, the equivalency between the SAR and the sum of OARs is independent of any kind of spatial placement (Šizling and Storch 2004). We can therefore use the sum of OARs to model the SAR without any limitations. The response variable of the OAR is the proportion of samples occupied (p), which is by definition the probability that a randomly selected sample (from the arena) is occupied by the focal

6

species. As introduced in the above paragraph, this probability can be computed from maps of species ranges as the ratio between the planar sizes of two areas: (a) the planar size of the area of the locations where sample plots would be *deemed occupied*; and (b) the planar size of the area of *all sample locations* within the arena (the areas delimited by thin lines in Fig. 1). Although the rules governing the geometric processes are independent from the choice of arena, the rules governing the biological processes may vary between focal arenas and taxa. The arena is then usually a "naturally" defined area such as a discrete habitat area, biome, island or continent (the area within the bold line in Fig. 1a).

The area where sample plots would be deemed occupied by a given species is here called the *effective range*. The effective range is larger than the actual spatial range of the species; how much it is larger depends on σ. This is because some sample plots have their centres located outside the species' range but would be deemed occupied because they overlap with the range (Fig 1b). Therefore, circular sample plots enlarge the effective range by the radius of a sample plot (ρ; $\sigma = \pi \cdot \rho^2$) all along the outline of the range. Similarly, square sample plots enlarge the effective range by approximately half of their edge along the outline of the range. In the approach used by Storch et al. (2012), the area where sampling plots can be placed entirely within the arena is smaller than the actual arena (depending on σ), and we term this here the *effective arena*. The effective arena is smaller than the arena by a band along its margins that is the width of ρ.

The area of the effective arena decreases and area of the effective range increases with increasing σ (see black insets in Fig 1). The rates of the increase and decrease, however, vary depending on the area of the gaps (the patches within a range that are not occupied by the species) and the part of the effective range that is not sampled because the sample plots would be located outside the effective arena. Here we call the part of the effective range that would be outside the effective arena FAE, for it

is caused by Finite-Area Effect (Šizling and Storch 2004, but see also Fig. 2c). FAE is therefore a variable in planar units, $\mathrm{FAE}(\sigma)$ is a function of σ, and finite-area effect is the phenomenon.

Some SAR-makers allow their sample plots to exceed to some extent the edges of the arena. In this case, the effective arena is larger than the effective arena as modelled here, depending on the particular design. These sampling designs vary in the width of the band that runs around the arena and that makes the effective arena different from the arena. For these sampling designs, we would only modify the expression for the bandwidth in our equations. Everything else, including the effective range and finite-area effect, holds unchanged in our theory.

Effective Arena

The above section describes how the arena and the effective arena differ from each other by a band, with a width of ρ (i.e., radius of sample plot; see Tab. 1 for the system of notation), that runs along the edge of the arena (Figs 1a and 2a-c). This is because all sample plots that have their centres (i.e., midpoints) closer to the edge of the arena than ρ would also sample areas outside the arena. The areas of arena (A) and effective arena (A_e) thus scale as

$$A_e \cong A - c' \cdot \rho \tag{2}$$

where c' (Fig. 1) stands for the length of a line that is parallel to the edge of the arena at distance $\rho/2$ from its edge. The parameter c' is smaller than the circumference of the arena (c) so that

$$c' \cong c - \pi \cdot \rho. \tag{3}$$

This is because the edge of an arena makes a closed curve. Therefore the shortening of the circumference of the arena follows the shortening of a circle with a radius that is shortened by $\rho/2$. This

shortening is independent from the radius of the circle and thus it is independent from the area of the

arena (for detailed reasoning see SI-1).

A planar measure of a shape generally scales with the second power of its linear size.

Consequently, the circumference of the arena, c, scales with $\sqrt{A}$ and shape specific parameter κ. Hence

$$c \cong 2 \cdot \sqrt{\kappa \cdot A}. \tag{4}$$

In the case of a circular arena $\kappa = \pi$ and in the case of a square shaped arena $\kappa = 4$.

Combining Eqs 2-4, we get an equation that returns A_e, which is

$$A_e \cong A - \left(2 \cdot \sqrt{\kappa \cdot A} - \pi \cdot \rho\right) \cdot \rho = \left(\sqrt{A} - \sqrt{\pi} \cdot \rho\right)^2 + 2 \cdot \sqrt{A} \cdot \rho \cdot \left(\sqrt{\pi} - \sqrt{\kappa}\right). \tag{5}$$

Hereafter we assume that the expression '$2 \cdot \sqrt{A} \cdot \rho \cdot \left(\sqrt{\pi} - \sqrt{\kappa}\right)$' in Eq. 5 is significantly smaller than

$\left(\sqrt{A} - \sqrt{\pi} \cdot \rho\right)^2$, and therefore we can omit it from the equation. Applying this assumption, the area of

the effective arena can be estimated as

$$A_e \cong \left(\sqrt{A} - \sqrt{\sigma}\right)^2 \tag{6}$$

where $\sigma = \pi \cdot \rho^2$ ('σ' and 'scale' will be hereafter treated as synonyms, in accord with the mathematical

and part of the macroecological literature, e.g., Kunin 1998, Harte et al. 1999, Lennon et al. 2002). For

practical reasons, we will hereafter treat A as the area of maximum σ (i.e. σ_{max}) that can be plotted

within the arena (i.e. we redefine $A := \sigma_{max}$). The rationale behind this is that A_e approaches zero as σ

approaches its maximum, which is smaller than A (i.e. $\sigma_{max} < A$) in most of real cases. However, in our

approximate equation for effective arena size (Eq. 6), A_e approaches zero as σ approaches the area of

the whole arena ($\sigma \to A$). However, σ_{max} could only reach the A ($\sigma_{max} = A$) if the arena and sample

plot were similar in shape (e.g. if both were circles or both were squares). In conclusion, our estimate of

A_e would be inaccurately large in real cases, if we did not approach A using σ_{max}. The approach to A_e

using Eq. 6 is useful particularly when we have no detailed information on the arena.

Our approach works poorly when the arena has a fractal-like shape with narrow peninsulas, as in

Europe with the peninsulas of Greece and Italy. In the case of Europe, a sample plot with a diameter (i.e.

2ρ) larger than the width of Italy cannot be plotted within this peninsula. The arena thus becomes

Europe without the Apennine peninsula when σ is large, and it is Europe with the peninsula when σ is

small. Nevertheless, a map of the arena together with the shapes and sizes of the sample plots should

be sufficient to estimate accurately the relationships between A, A_e, and σ.

Effective Range

The effective range of a species is a range that is enlarged by a band the width of ρ (the range including

this band is hereafter referred to as Outline_e) but without FAE (i.e. Finite-Area Effect; Fig. 2c) and

without effective gaps in the range ($\text{Gap}_{e,j}$ stands hereafter for the jth effective gap; Fig. 1b). Effective

gaps, similar to the effective arena, are defined as the areas where sampling plots can be placed entirely

within the gaps (Fig. 1b). The equation for the area of the ith effective range ($a_{e,i}$) obeys

$$a_{e,i} = \text{Outline}_{e,i} - \sum_{j=1}^{Ngi} \text{Gap}_{e,i,j} - \text{FAE}_i \qquad (7)$$

where $N_{g,i}$ is the number of effective gaps within the *i*-th range.

The areas of Outline_e, and Gap_e can also be computed in the same way as that of the effective

arena. So

$$\text{Outline}_{e,i} = a_i + c'_{r,i} \cdot \rho = a_i + 2 \cdot \sqrt{\kappa_{a,i} \cdot a_i} \cdot \rho + \pi \cdot \rho^2 = a_i + 2 \cdot \sqrt{\frac{\kappa_{a,i}}{\pi}} \cdot \sqrt{a_i} \cdot \sqrt{\sigma} + \sigma \tag{8}$$

where a_i is the area within the outline of the ith range (i.e. the area of a range as if the range had no

gaps), and $c'_{r,i}$ is the length of the line that runs around the range at a distance of $\rho/2$ outside the range

outline (Fig. 1b). Finally

$$\text{Gap}_{e,i,j} = g_{i,j} - c'_{g,i} \cdot \rho = g_{i,j} - 2 \cdot \sqrt{\kappa_{g,i,j} \cdot g_{i,j}} \cdot \rho + \pi \cdot \rho^2 \tag{9}$$

$$= g_{i,j} - 2 \cdot \sqrt{\frac{\kappa_{g,i,j}}{\pi}} \cdot \sqrt{g_{i,j}} \cdot \sqrt{\sigma} + \sigma$$

where $g_{i,j}$ is the area of the jth gap within the ith range, and $c'_{g,i}$ is the length of a line running around the

gap at the distance $\rho/2$ inside the gap outline (Fig 1b). Moreover, $\kappa_{a,i}$ and $\kappa_{g,i,j}$ are analogous to the κ

in Eq. 4, i.e., the kappas are the shape specific constants of the ith range and the jth gap.

The sum in Eq. 7 runs across all the gaps in the focal range. However, some gaps may be so small

or so narrow that all sample plots located within them sample only part of the range. The effective gaps

then become non-existent and consequently no gap area will be subtracted from the range. It is

therefore practical to run the summation in Eq. 7 only across the effective gaps at the scale in question

(or to put $\text{Gap}_{e,i,j} = 0$ where $\text{Gap}_{e,i,j}$ would be ≤ 0).

The FAE of the ith species as defined earlier in the text (i.e. the area that would have belonged to

the effective range had it not been outside the effective arena; Figs. 2c-d and SI-2) is a function of

distance between the edge of the arena and the range. FAE depends on whether the closest edge of the

arena is convex or concave. Nevertheless, having no information on the exact location of the range

within the arena, we should assume that the closest edge of the arena is a part of a circle with the same

area as the arena. This is because any deviation of the edge of the arena from the circle would need

information about the arena shape and range location. The circles are therefore our null expectations

about the arena and ranges. Our null expectation about *FAE* is then the area of the crescent-shaped part

of the effective range that is cut off by the edge of the effective arena (see the missing part of the circle

that represents the effective range in Fig. 2c and SI-2). It follows that *FAE* is zero whenever the effective

range lies entirely within the effective arena (Figs. 2a,b), and the effective range fills the effective arena

when the effective range would be so large that it would entirely encompass the effective arena (Fig. 2d,

SI-2). In between these two extremes the *FAE* is a function of σ (with fixed A, a_i, and d_i that stands for

the distance between the edge of the arena and the range centre; Fig. 2c), and it obeys

$$\text{FAE}_i(\sigma) = r_{e,i}^2 \text{acos}\left(\frac{R_e^2 - r_{e,i}^2 - (R - d_i)^2}{2(R - d_i)r_{e,i}}\right) - R_e^2 \text{acos}\left(\frac{R_e^2 - r_{e,i}^2 + (R - d_i)^2}{2(R - d_i)R_e}\right) \tag{10}$$

$$+ (R - d_i)\left(R_e^2 - \left(\frac{R_e^2 - r_{e,i}^2 + (R - d_i)^2}{2(R - d_i)}\right)^2\right)^{0.5}$$

(see SI-2 for derivation). Equation 10 is not explicitly dependent on σ. However, the dependence of *FAE*

on σ becomes obvious if we replace the radius of the effective arena, R_e, with $\sqrt{A/\pi} - \sqrt{\sigma/\pi}$, and the

radius of the effective range, $r_{e,i}$, with $\sqrt{a_i/\pi} + \sqrt{\sigma/\pi}$. The expressions $\sqrt{A/\pi}$, and $\sqrt{a_i/\pi}$ represent

virtual radii of the arena (i.e. R in the Eq. 10) and the range, respectively; and $\sqrt{\sigma/\pi}$ stands for ρ. We

replaced the radii with functions of areas, because 'real' arenas and ranges are not generally circular.

The radii are therefore only virtual values in case of real (non-circular) arenas and ranges. This

assumption is appropriate because *FAE* depends only on the part of the arena edge that is nearest to the

range in question, while the opposite side of the arena is irrelevant.

Occupancy–Area Relationship (OAR)

Defining the effective range and the effective arena simplifies our definition of the OAR, which can now

be expressed as:

$$p_i(\sigma) = \frac{a_{e,i}(\sigma)}{A_e(\sigma)} \tag{11}$$

where p_i is the probability of occupancy by the ith species. The equation to compute A_e holds across all

scales, whereas the equation to compute $a_{e,i}$ (Eq. 7) is scale dependent. As a consequence, OARs are

moulded by different processes over different ranges of scale. At fine scales, when σ becomes small,

the effective range approaches the actual range, and so $a_{e,i}$ has to be reduced by subtracting Gap_e of all

the gaps that are bigger than σ (Fig 2a). At progressively coarser scales, samples increasingly encompass

all the gaps ($\sigma \geq \mathrm{Gap}_{e,i,j}; \ \forall j$), and the effective range appears compact with no gaps (Fig 2b).

Eventually, the effective ranges will extend beyond the effective arena ($\mathrm{FAE}_i > 0$), reducing the

occupancy by finite-area effect (Fig 2c), until ultimately the occupancy "saturates" at the coarsest scales,

where it becomes impossible for a sample not to include the species (Fig 2d; see 'area of saturation' in

Šizling and Storch 2004). Each process acts over a range of scales, and each species enters these phases

at a different σ. The overall SAR depends on the fraction of species in each state at a given scale.

Various mechanisms that mould OARs at various scales are captured by different equations. At the

finest scales, where the range already appears as a solid area but some effective gaps still exist (Fig. 2a),

the equation for OAR obeys

$$p_i(\sigma) \approx \frac{\mathrm{Outline}_{e,i} - \sum_{j=1}^{Ngi} \mathrm{Gap}_{e,i,j}}{\left(\sqrt{A} - \sqrt{\sigma}\right)^2}. \tag{12}$$

13

(see eqs 8 and 9 for the Outline_e and Gap_e). At a coarser scale (meaning the OAR is generated from larger sample plots), where no gap is larger than σ (Fig. 2b)

$$p_i(\sigma) \approx \frac{\text{Outline}_{e,i}}{\left(\sqrt{A} - \sqrt{\sigma}\right)^2}. \tag{13}$$

At a still coarser scale, the effective ranges will meet the edge of the effective arena (Fig. 2c)

$$p_i(\sigma) \approx \frac{\text{Outline}_{e,i} - \text{FAE}_i(\sigma)}{\left(\sqrt{A} - \sqrt{\sigma}\right)^2}. \tag{14}$$

At such scales, some effective gaps could exist and if so, $\text{Gap}_{e,i,j}$ would have had to be subtracted from the nominator of Eq. 14. In this case (i.e., coexistence of non-zero FAE and $\text{Gap}_{e,i,j}$), however, the scale captured by Eq. 13 disappears. Finally, when the sample plot becomes so large that the effective range entirely fills the effective arena (Fig. 2d), the OAR obeys

$$p_i(\sigma) = 1. \tag{15}$$

This means that σ has reached the area-of-saturation (Šizling and Storch 2004).

Species–Area Relationship (SAR)

Each SAR can be generated as the sum of OARs. Species richness, *S*, as a function of scale then follows

$$S(\sigma) = \sum_{i=1}^{Stot} p_i(\sigma) \tag{16}$$

where S_{tot} stands for the total species richness of the taxonomic group in question within the arena. A closer look at the definition of p (Eq. 11) shows that the summation is needed only for effective ranges (i.e. the numerator), because only effective ranges vary between the species, whereas the effective arena (i.e. the denominator) is species independent. The functions Outline_e and Gap_e (Eqs. 8 and 9,

respectively) are both combined from three additive terms that are (i) independent of σ, (ii) dependent on $\sqrt{\sigma}$, and (iii) dependent on σ. The relationship between *FAE* and σ is complex. As a result, the gaps and the area within the range outline scales in a different way than *FAE* (i.e., it follows different equations), which needs to be considered when rescaling SARs. It is therefore useful to differentiate between summations of the Outline_es and Gap_es on one hand, and FAEs on the other. Species richness then scales as

$$S(\sigma) \approx \frac{C_1\sigma + C_2\sqrt{\sigma} + C_3 - \text{FAE}_{tot}(\sigma)}{\left(\sqrt{A} - \sqrt{\sigma}\right)^2}, \tag{17}$$

which is the mathematical formula of geometric processes in the SAR. Its coefficients C_{1-3} are modified by biological processes affecting the SAR. In this formula, the coefficients C substitute for

$$C_1 = \sum_{i=1}^{Stot} 1 - \sum_{m=1}^{Ng(\sigma)} 1 = S_{tot} - N_g(\sigma) \tag{18}$$

where N_g is the number of effective gaps across all species. N_g is a function of scale as sample-plots increasingly encompass more gaps and effective gaps therefore vanish with increasing σ. We can replace the pair of index variables that indicate the species range (i) and the gap within the range (j) with one index m. The reason for this is that the summation runs across all effective gaps whatever the species identity. Then

$$C_2 = \frac{2}{\sqrt{\pi}}\left(\sum_{i=1}^{Stot} \sqrt{\kappa_{a,i} \cdot a_i} + \sum_{m=1}^{Ng(\sigma)} \sqrt{\kappa_{g,m} \cdot g_m}\right), \tag{19}$$

$$C_3 = \sum_{i=1}^{Stot} a_i - \sum_{m=1}^{Ng(\sigma)} g_m \tag{20}$$

(a result of the summation of the Eqs. 8 and 9) and

15

$$\text{FAE}_{tot}(\sigma) = \sum_{i=1}^{Stot} \text{FAE}_i(\sigma). \tag{21}$$

Because N_g decreases as the effective gaps vanish, N_g is a function of scale and so are the second summations in Eqs. 18-20. Neither N_g nor the second summations, however, change in the intervals of σ that are between two subsequent points at which effective gaps vanish (i.e., at which N_g decreases by one). In these intervals, the coefficients C hold constant; and these intervals are large mainly at (i) large scales where no effective gap exists and (ii) at small scales where all gaps contribute to the summation.

In sum, the mathematical expression for species richness as a function of scale follows a fraction of two polynomial-like functions with three orders (one of them non-integer) that are 1, 0.5 and 0. The coefficients of the polynomial-like function in the numerator scale with (i) the difference between numbers of ranges and effective gaps, (ii) with the total circumference of ranges including their gaps $(2 \cdot \sqrt{\kappa_{a,i} \cdot a_i}$, and $2 \cdot \sqrt{\kappa_{g,m} \cdot g_m}$ are the circumferences of ith range and m^{th} gap; see Eqs. 4,8 and 9); and (iii) with the area of all ranges without the areas of their gaps. Because an increase in σ results in a decrease in the number of effective gaps, the coefficients C_{1-3} hold only for each particular scale, and they discontinuously change at each scale break, as the effective gaps disappear. At large scales, S decreases by the sum of FAEs, which quantifies the finite-area effects and that are complex functions of σ and d_is.

Rescaling of SAR

In mathematical notation, the rescaling by Storch et al. (2012) (Eq. 1) is expressed by replacing σ with '$\bar{a} \cdot \sigma^*$' and dividing the value $S(\bar{a} \cdot \sigma^*)$ with $S(\bar{a})$ where the asterisk labels the rescaled variables and $\bar{a}$ is the mean area across all ranges. This rescaling modifies the formula of SAR (Eq. 17) into

$$S^*(\sigma^*) \approx \frac{C_1 \cdot \bar{a} \cdot \sigma^* + C_2 \cdot \sqrt{\bar{a}} \cdot \sqrt{\sigma^*} + C_3 - \mathrm{FAE}_{tot}(\bar{a} \cdot \sigma^*)}{\mathrm{S}(\bar{a}) \cdot \left(\sqrt{A} - \sqrt{\bar{a}} \cdot \sqrt{\sigma^*}\right)^2}. \tag{22}$$

Eq. 22 shows that a SAR rescaled by Storch et al. (2012) depends on the function for finite-area effect (FAE) and reduction of the effective arena with expanding scale (bracket in the denominator). Both these effects are imposed by the edge of the arena and are results of geometric processes. Therefore, we hereafter term them geometric edge effects or simply the *focal edge effects*. The biological intuition for the existence of a universal SAR that is standardized by $\bar{a}$ and $S(\bar{a})$ (e.g., Storch et al. 2012, Lazarina et al. 2013) does not consider these focal edge effects. We can, however, rearrange Eq. 22 in order to capture the ideal SAR (i.e., the SAR that might follow biological intuition for collapse) into square brackets. Eq. 22 then, after being rearranged, yields

$$S^*(\sigma^*) \approx \frac{A \cdot \bar{a}}{S(\bar{a}) \cdot \left(\sqrt{A} - \sqrt{\bar{a}} \cdot \sqrt{\sigma^*}\right)^2} \left[\frac{\sigma^* \cdot C_1 + \sqrt{\sigma^*} \cdot C_2/\sqrt{\bar{a}} + C_3/\bar{a}}{A}\right] \tag{23}$$

$$- \frac{A \cdot \bar{a}}{S(\bar{a}) \cdot \left(\sqrt{A} - \sqrt{\bar{a}} \cdot \sqrt{\sigma^*}\right)^2} \cdot \frac{\mathrm{FAE}_{tot}(\bar{a} \cdot \sigma^*)}{A \cdot \bar{a}}.$$

The term within the square bracket in Eq. 23 (let us call it $S'(\sigma^*)$) is a virtual SAR, i.e., a SAR without focal edge effects. $S'(\sigma^*)$ is a virtual value because focal edge effects cannot be removed by any sampling design, but only by means of mathematical analysis. The virtual SAR is simply the sum across all effective ranges without subtracting *FAE* and divided by the area of the whole arena (i.e. A). In addition, coefficients are standardized by $\bar{a}$. To be precise, the coefficient C_1 is not defined spatially, and therefore it is not standardized by $\bar{a}$; C_2 is a circumference of ranges (a linear measure), and therefore it is standardized by $\sqrt{\bar{a}}$. Finally, C_3 and *FAE* are planar measures across all ranges and effective gaps, and therefore they are standardized by $\bar{a}$.

When substituting $S'(\sigma^*)$ for the square bracketed term and simplifying the fractions, Eq. 23 turns into

$$S^*(\sigma^*) = \frac{1}{S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{\sigma^*/A}\right)^2} \cdot S'(\sigma^*) - \frac{\text{FAE}'_{tot}(\bar{a} \cdot \sigma^*)}{S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{\sigma^*/A}\right)^2} \tag{24}$$

where

$$\text{FAE}'_{tot}(\bar{a} \cdot \sigma^*) = \frac{\text{FAE}_{tot}(\bar{a} \cdot \sigma^*)}{A \cdot \bar{a}} \tag{25}$$

is a standardized function of the total finite area effect.

$S'(\sigma^*)$ is not standardized to the differences in species pool (S_{tot}) between different arenas.

Storch et al. (2012) used the value of $S(\bar{a})$ to standardize their rescaled SAR to the species pool (Eq. 1).

However, this value cannot standardize our $S'(\sigma^*)$ to the variation in species pools. This is because

$S'(\sigma^*)$ is a virtual value (i.e. it cannot be observed and has to be computed) while $S(\bar{a})$ is the 'real'

species richness, and the virtual value cannot be standardized by a 'real' value. The value of $S'(\sigma^*)$

therefore has to be divided by $S'(1)$, which is the virtual species richness at scale $\sigma = \bar{a}$ (i.e., $\sigma^* = 1$).

The value of $S'(1)$ can be extracted from Eq. 24 by replacing σ^* with one, which gives

$$S^*(1) = 1 \approx \frac{1}{S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{1/A}\right)^2} \cdot S'(1) - \frac{\text{FAE}'_{tot}(\bar{a})}{S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{1/A}\right)^2}. \tag{26}$$

($S^*(1) = 1$ is a simple consequence of the rescaling by Storch et al. (2012), Eq. 1, because $S^*(1) =$

$S(\bar{a} \cdot 1)/S(\bar{a})$.) By solving Eq. 26 with respect to $S'(1)$, it follows that

$$S'(1) \approx S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{1/A}\right)^2 + \text{FAE}'_{tot}(\bar{a}). \tag{27}$$

The $S'(\sigma^*)$ standardized by $S'(1)$ becomes

$$S^{**}(\sigma^*) = \frac{S'(\sigma^*)}{S'(1)}, \tag{28}$$

which is the rescaled SAR without focal edge effects, standardized to the variability in species pool. The S^{**} may, therefore, aspire to be the universal SAR to which all rescaled SARs collapse, and it is related to $S^*(\sigma^*)$, as

$$S^*(\sigma^*) = \frac{S^{**}(\sigma^*) \cdot S'(1) - \text{FAE}'_{tot}(\bar{a} \cdot \sigma^*)}{S(\bar{a}) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{\sigma^*/A}\right)^2}. \tag{29}$$

Eq. 29 is a combination of Eqs. 24 and 28. The relationship between the here rescaled SAR, $S^{**}(\sigma^*)$, and the observed SAR ($S(\sigma)$) then follows from Eq. 29 and $S^*(\sigma^*) = S(\sigma)/S(\bar{a})$ (see Eq. 1), and it obeys

$$S^{**}(\sigma^*) = \frac{S(\sigma) \cdot \left(\sqrt{1/\bar{a}} - \sqrt{\sigma^*/A}\right)^2 + \text{FAE}'_{tot}(\bar{a} \cdot \sigma^*)}{S'(1)}. \tag{30}$$

For practical purposes, Eq. 29 can serve to extract expected species richness at σ from (i) universal $S^{**}(\sigma^*)$, (ii) taxon and arena specific $\bar{a}$, and (iii) arena specific A. Eq. 30, in turn, can serve to test whether the observed SAR (i.e., $S(\sigma)$) collapses to $S^{**}(\sigma^*)$. Because we usually do not have details on FAE'_{tot}, we recommend estimating its value from Eq. 10 by replacing a_i with $\bar{a}$ and d_i with $(2/3) \cdot \sqrt{A/\pi}$. The rationale behind this is that $\bar{a}$ is the mean value across all a_i, and $(2/3) \cdot \sqrt{A/\pi}$ is the mean value across all d_i given a zero expectation (i.e., given that the arena is a circle, and the range centres are evenly distributed). This gives an estimation of the average FAE for one 'average' species. The estimation of FAE'_{tot} therefore scales with the product of this average FAE and $S_{tot}/(\bar{a} \cdot A)$.

Biological processes

Eq. 30 is a formula for our rescaled SAR, $S^{**}(\sigma^*)$, that eliminates focal-edge effects from the observed SAR, $S(\sigma)$. Eq. 30 therefore carries no information on the biological processes behind the SAR. This

19

information is carried by the parameters C. $S^{**}(\sigma^*)$ in terms of parameters C arises from the term in

square brackets in Eq. 23,

$$S'(\sigma^*) = \frac{\sigma^* \cdot C_1 + \sqrt{\sigma^*} \cdot C_2/\sqrt{\bar{a}} + C_3/\bar{a}}{A}, \tag{31}$$

and the standardization by $S'(1)$, which can be expressed in terms of the coefficients C as

$$S'(1) = \frac{C_1 + C_2/\sqrt{\bar{a}} + C_3/\bar{a}}{A} \tag{32}$$

(Eq. 32 follows from Eq. 31). The formula for the here rescaled SAR that captures the biological

processes therefore follows

$$S^{**}(\sigma^*) = \frac{S'(\sigma^*)}{S'(1)} = \sigma^* \cdot C_1' + \sqrt{\sigma^*} \cdot C_2' + C_3', \tag{33}$$

where C_{1-3}' are the standardized coefficients ($C_1' = C_1/(C_1 + C_2/\sqrt{\bar{a}} + C_3/\bar{a})$,

$C_2' = (C_2/\sqrt{\bar{a}})/(C_1 + C_2/\sqrt{\bar{a}} + C_3/\bar{a})$, and $C_3' = (C_3/\bar{a})/(C_1 + C_2/\sqrt{\bar{a}} + C_3/\bar{a})$; note that $C_1' + C_2' +$

$C_3' = 1$). Apparently, $S^{**}(\sigma^*)$ of two assemblages A and B are equal to each other (i.e., collapse) iff

(meaning 'if and only if') the standardized coefficients are equal to each other, i.e., iff

$$C_{1A}' = C_{1B}' \,\&\, C_{2A}' = C_{2B}' \,\&\, C_{3A}' = C_{3B}'. \tag{34}$$

Data test

The actual rescaled SAR (S^*) and the ideal rescaled SAR (S^{**}) are related by the term $\sqrt{1/\bar{a}}$ (Eq. 29),

given that FAE is constant or zero, which happens mainly at small scales. Storch et al. (2012) and

Lazarina et al. (2013) both show patterns that reveal different slopes for each rescaled SAR. These SARs

all intersect at point $[\sigma^*, S^*] = [1,1]$ (a simple consequence of Eq. 1 where we put $\sigma^* = 1$), making a

collection of rescaled SARs, which resembles a hand-fan with a set of rays radiating from a point of

20

attachment, and so we call it hereafter the 'hand-fan of SARs'. Our theory predicts that the hand-fan of

SARs is driven by $\bar{a}$. We used the Wolfram Computational Engine, which is accessible at

www.**wolframalpha**.com, to compute the slopes of logarithmically transformed S^*. These slopes follow

$$Slope = \frac{\partial \ln S^*(\sigma^*)}{\partial \ln \sigma^*} \approx \frac{K_1\sqrt{\bar{a}} + K_2}{K_3\sqrt{\bar{a}} + K_4}. \tag{35}$$

Although Eq. 35 looks simple, the derivative, given by the computational engine, is a complex equation.

We are, however, focusing on the effect of $\bar{a}$ and so we can substitute all the variables in the complex

equation (except the $\bar{a}$) with four coefficients K. See SI-3 for the Ks expressed from previously

introduced parameters. In our test, we extracted the values of K_{1-4} by least square regression across

data adopted from Storch et al. (2012) and Lazarina et al. (2013). The coefficients K_{1-4} formally depend

on σ^*, but we ignore this dependency here.

Our prediction (Eq. 35) explains the data used by Lazarina et al. (2013) well (Fig. 3a; $p < 1 \cdot$

10^{-6}, $r^2 = 0.637$, N=165), where $\bar{a}$ was adopted from their Tab. 1 and slopes were extracted by

digitalizing their Figs 1a-b. However, we received a poorer fit of Eq. 35 to the data used in Storch et al.

(2012) (Fig. 3b). (The data were taken from the Figs 1,2 in Storch et al. 2012.) The reason for the poorer

fit in the latter case than in the former one is that the data used by Storch et al. (2012) are more

scattered around our prediction than the data by Lazarina et al. (2013). This is because the slopes of

SARs in double-logarithmic space naturally vary across scales (i.e., SARs are not perfectly power laws

with constant slopes in double-logarithmic space). Therefore, depending on the scales in focus, a variety

of slopes are attributed to each particular $\bar{a}$. This effect was, for the sake of simplicity, ignored when we

derived the Eq. 35 (SI-3). Because the SARs presented by Storch et al. (2012) deviate from the power-law

more than the SARs presented by Lazarina et al. (2013), the fit is poorer in the first case. Importantly,

however, the correlation between the slope and $\bar{a}$ is still apparent and significant (Fig. 3b; $p<1\cdot10^{-5}$,

$r^2 = 0.1136$, N=267). To support this reasoning, we approximated the SARs presented in Lazarina et al.

(2013) by exact power-laws (which cannot be done for the SARs by Storch et al. (2012) due to their curvilinearity). By doing so, we generated a unique slope for each single $\bar{a}$ and our prediction (Eq. 35) was met, almost perfectly (SI-4).

The above test is vital to see whether our model works. However, our model may be accurate, even if there is no observed collapse. This would happen if the biological intuition for the collapse failed and the biological processes modified the parameters C differently for each assemblage. The next test is therefore designed to show whether we should reject the assumed collapse or not. We used the SARs, $\bar{a}$s and $S(\bar{a})$s extracted from Storch et al. (2012) and Lazarina et al. (2013). We then rescaled σ in the same way as these authors and species richness using our Eq. 30 (the reader can also use our excel sheet 'SI-5: Step by step rescaling' to rescale any SAR using the observed $S(\sigma)$, $\bar{a}$, $S(\bar{a})$, and A).

We found an improvement in the assumed collapse for almost all SARs reported by Storch et al. (2012) and Lazarina et al. (2013) (compare Figs. 4ab). Only two datasets on woody plants of Mt. Holomontas adopted from Lazarina et al. (2013) failed to collapse (triangles in Fig. 4), meaning that the biological processes may work differently in these habitats. The collapse of S** was particularly noticeable for small arenas, where no collapse was identified by Lazarina et al. (2013).

Discussion

We have resolved the controversy between Storch et al. (2012) and Lazarina et al. (2013) in favour of the latter authors. We have shown that the problem demonstrated by Lazarina et al. (2013) affects not only small, but all scales. To put it in more technical terms, Storch et al. (2012) suggested rescaling by a mean range area ($\bar{a}$), which implies that the rescaled SAR should be independent of this parameter. Our

model has, however, revealed a correlation between the slope of rescaled SAR and $\bar{a}$, which refutes the hypotheses of the collapse proposed by Storch et al. (2012).

Although we agree with Lazarina et al. (2013) that the rescaling proposed by Storch et al. (2012) does not work, we disagree with the reasoning by Lazarina et al. (2013). Lazarina et al. (2013) stated that their SARs did not collapse and Storch et al.'s (2012) SARs did collapse because β-diversity is maximized and its effect on SAR shape is minimal for the arenas of continental size; and because, for small arenas, SARs did not collapse due to high variation in β-diversity as species ranges vary in their overlap. We argue that β-diversity drives the SAR equally for a variety of arenas and at all scales (Harte and Kinzig 1997; Koleff et al. 2003; Gaston et al. 2007; Tjørve and Tjørve 2008; Šizling et al. 2011). Moreover, we argue that SAR is independent of the species-range overlap because it is given by the simple sum of OARs (Coleman 1981), which is attributed to individual species, with no need for interaction between them. We find that the observed variance in rescaled SARs is caused by variation in $\bar{a}$ at all scales, in addition to the finite-area effect (FAE; Šizling and Storch 2004) at large scales. As proof of our reasoning, we demonstrated that our model produces a better collapse of rescaled SARs by controlling for these two effects.

Visually, the rescaled SARs in Storch et al. (2012) and Lazarina et al. (2013) have different slopes at each scale and intersect each other at one point, which makes a hand-fan of rescaled SARs. This hand-fan of rescaled SARs is much more prominent in Lazarina et al. (2013) than in Storch et al. (2012). This is because Storch et al. (2012) put more emphasis on the larger scales (Lazarina et al. 2013) where the logarithmic function is increasingly flattened, which results in seemingly more similar values at large scales and more different values at small scales. At small scales, where the logarithmic function can be approximated by a line, the hand-fan of rescaled SARs is equally visible in both reports (see our Fig. 4 and capture to Fig. 1 in Storch et al. 2012). This effect veils a bias in Storch et al.'s (2012) rescaled

estimates. This bias is of about 0.07 from the universal curve, equivalent to 17% (i.e., $100(10^{0.07} - 1)$%) of continental diversity.

The collapse of the newly rescaled SARs presented here is better than that demonstrated by Storch et al. (2012) and Lazarina et al. (2013). This suggests that the intuition of Storch et al. (2012) for the collapse may have worked, but that no collapse was found in Lazarina et al. (2013) or in Storch et al. (2012), because these authors ignored the geometric effects that were imposed by the edge of the arena. Lazarina et al. (2013, p. 965) summarized the reasons for the expected collapse in terms of 'maximized species turnover' and 'universal properties of Species-Abundance Distribution (SAD)' at large scales (Rosindell and Cornell 2013, for the mechanism see Šizling et al. 2009a,b, Kůrka et al. 2010). Our analysis identified the specific conditions that will lead to a collapse. The rescaled SAR, in our approach, is a 'polynomial' function of rescaled σ (σ^*) with two integer and one non-integer orders (1; 0.5; 0) (Eq. 33). Its coefficients C_1', C_2' and C_3' are moulded by biological processes and two rescaled SARs collapse onto each other iff these coefficients equal to each other (Eq. 34). The rationale behind this is that two polynomials only collapse if their respective coefficients are identical. We therefore infer four formal conditions of the collapse: (i) equality of C_1's suggests that standardized differences between the numbers of ranges and their gaps are equal across the assemblages with collapsing SARs; (ii) equality of C_2's suggests that standardized sums of perimeters (including perimeters of gaps) across all ranges are equal across the assemblages with collapsing SARs; and (iii) equality of C_3's suggests that standardized sums of range areas (without the gaps) are equal across the assemblages with collapsing SARs. In addition, the coefficients C_{1-3}' vary as the effective gaps (i.e., the gaps visible at each scale) vanish with increasing scale, and therefore these three conditions on collapse have to be satisfied at each scale. The scales are separated by the persistence of effective gaps. An effective gap vanishes whenever the sample plot encompasses the gap. Hence, the fourth condition (iv) concerns standardized frequency distribution of gap areas, and gap circumferences. These frequency distributions have to be identical

24

across all assemblages and arenas with SARs collapsing to the universal curve. We do not argue that our

and Lazarina et al.'s (2013) reasons for the collapse are in conflict. However, the link between the

conditions that would lead to a collapse by Lazarina et al. (2013) and the conditions 'i-iv' identified by

our approach needs detailed examination.

The collapse demonstrated here is not entirely general. Datasets on woody plants of Mt.

Holomontas (HOL$_I$ and HOL$_{II}$ in Lazarina et al. 2013) failed to collapse (Fig. 4b), indicating either that the

conditions for a collapse from the paragraph above were not met, or that our simplification of the A_e

computation (Eq. 6) failed in this case. Although the SARs based on the HOL$_I$ and HOL$_{II}$ data did not

collapse on the same curve as did the other 24 SARs, these two SARs collapsed on each other (compare

open triangles in Fig. 4a and 4b) indicating similar biological processes in the region.

We have improved the collapse assumed by Storch et al. (2012) by subtracting FAE from each

range. FAE is a function of σ and distance between the focal range and the edge of the arena, which is

the coastline in Storch et al. (2012) and the edge of the extent of sampling in Lazarina et al. (2013).

Because neither Storch et al. (2012) nor Lazarina et al. (2013) controlled their collapse for FAE, the

quality of their collapse was dependent on the finite-area effect. This was the reason why the quality of

the collapse depended on the range locations in Storch et al. (2012), and why these authors

subsequently concluded that the ranges of assemblages, with collapsing SARs, would have had to be

regularly distributed within the arena. Our analysis shows that the assumption of regular distribution of

ranges is too strict. The necessary (but not the only) condition for collapse by Storch et al. (2012) would

be equal frequency distribution of standardized FAEs. To get equal distribution of standardized FAEs

between arenas or various taxa, we *do not need* equal spatial distributions of ranges within various

arenas and for different taxa. To get equal distribution of standardized FAEs we *do need* equal frequency

distribution of distances between ranges and the edge of the arena within various arenas and for

different taxa. The validity of our condition is also obvious from the fact that there is no need to involve interactions between species ranges into the computation of SAR from species occupancies (Eq. 16; He and Legendre 2002). This means that SAR is independent of the degree of overlap in species ranges, but it depends on distances of species ranges from the edge of the arena. If the arena was circular, then the SAR would not change were we to rotate each species' range-map independently: keeping the occupancy and position of each species relative to the arena's edge constant, while changing the degree of overlap with other species. Indeed, this mechanism caused that Storch et al. (2012) demonstrated a better collapse with real data than with simulations where these authors manipulated the frequency distribution of the distances between range and arena. However, the regular distribution of ranges within a continent that was assumed by Storch at al. (2012) is unlikely. There are multiple reasons for irregular distribution of species ranges within an arena. In addition to species-richness variation along several environmental and geographical gradients (e.g., Currie 1991, Willig et al. 2003, Currie et al. 2004, Hillebrand 2004, Brown 2014, Mannion et al. 2014), coastlines and edges of biomes will truncate ranges that otherwise would spread beyond the actual continent or biome edge (Šizling et al. 2009c). Moreover, species with small ranges (that are truncated) are more likely to become extinct than the species with large ranges (Johnson 1998) with their midpoints that are inevitably located closer to the centres of the continents or biomes (Colwell and Hurtt 1994).

Whatever the success of the model predictions presented here, we are still far from a solution for up-scaling. The relationships between our rescaled SAR, mean range size and FAE, as presented here, can certainly be used for a more efficient rescaling (resulting in better SAR collapse) given the here described biological conditions for collapse. Our rescaling as well as the one proposed by Storch et al. (2012) require accurate estimates of the mean range size. Nevertheless, there is always a huge number of rare species (e.g., Gregory 2000), which typically causes us to overestimate the mean range size, unless we have a detailed map of the whole arena including the ranges of all rare species. With a

detailed map, however, we neither need an up-scaling nor the rescaling. In conclusion, the quest for an effective tool to up-scale species richness from details on local assemblages still continues.

Summary

We split the processes that underlie SARs between biological processes and geometric processes (Eq. 17). Then we derived a function for the SAR that incorporates only the geometric processes while the biological processes are captured by the parameters of the function. Such a function is expressed as the ratio between the sum of *effective* ranges areas across all species and the area of the *effective* arena. The term "effective" relates to how large the range or the arena appears at the level of resolution given by the area of the sample plot (σ). A larger σ causes larger effective range and smaller arena. In the function for geometric processes (Eq. 17), we isolated the expression for the SAR without the geometric effects (square brackets in Eq. 23). We then rescaled this SAR as suggested by Storch et al. (2012). Because we did not rescale the observed SAR but the SAR that was purified from geometric effects, we achieved a better collapse than was demonstrated by Storch et al. (2012) and Lazarina et al. (2013). The relationships between the observed SARs ($S(\sigma)$), the SARs rescaled in accord with Storch et al. (2012) ($S^*(\sigma^*)$), and the here rescaled SARs ($S^{**}(\sigma^*)$) are given by Eqs. 29 and 30. The step-by-step rescaling is found in the Excel sheet in the supplementary information (SI-5). The reader can use this Excel sheet for user-friendly rescaling of their own data.

Acknowledgements

References

Allen, A. P. and White, E. P. 2003. Effects of range size on species–area relationships. – Evol. Ecol. Res. 5: 493–499.

Brown, J. H. 2014. Why are there so many species in the tropics? – J. Biogeogr. 41: 8–22.

Coleman, D. B. 1981. On random placement and species–area relations. – Math. Biosci. 54: 191–215.

Colwell, R. K. and Hurtt, G. C. 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. – Am. Nat. 144: 570–595.

Currie, D. J. 1991. Energy and large-scale patterns of animal–and plant–species richness. – Am. Nat. 137: 27–49.

Currie, D. J. et al. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. – Ecol. Lett. 7: 1121–1134.

Gaston, K. J. et al. 2007. The scaling of spatial turnover: pruning the thicket. – In: Storch D. et al. (eds), Scaling biodiversity. Cambridge Univ. Press, pp. 181–214.

Gregory, R. D. 2000. Abundance patterns of European breeding birds. – Ecography 23: 201–208.

Harte, J. et al. 2009. Biodiversity scales from plots to biomes with a universal species–area curve. – Ecol. Lett. 12: 789–797.

Harte, J. and Kinzig, A. 1997. On the implications of species–area relationships for endemism, spatial turnover, and foodweb patterns. – Oikos 80: 417–427.

Harte, J. et al. 1999. Self-similarity in the distribution and abundance of species. – Science: 284, 334–336.

He, F. L. and Legendre, P. 2002. Species diversity patterns derived from species–area models. – Ecology 85: 1185–1198.

Hillebrand, H. 2004. On the generality of the latitudinal diversity gradient. – Am. Nat. 163: 192–211.

Johnson, C. N. 1998. Species extinction and the relationship between distribution and abundance. – Nature 394: 272–274.

Koleff, P. et al. 2003. Measuring beta diversity for presence-absence data. – J. Anim. Ecol. 72: 367–382.

Kunin, W. E. 1998. Extrapolating species abundances across spatial scales. – Science 281: 1513–1515.

Kůrka, P. et al. 2010. Analytical evidence for scale-invariance in the shape of species abundance distributions. – Math. Biosci. 223: 151–159.

Lazarina, M. et al. 2013. Does the universality of the species–area relationship apply to smaller scales and across taxonomic groups? – Ecography 36: 965–970.

Lennon, J. J. et al. 2002. Fractal species distributions do not produce power-law species–area relationships. – Oikos 97: 378–386.

Mannion, P. D. et al. 2014. The latitudinal biodiversity gradient through deep time. – TREE 29: 42–50.

Muriel, N-N. and Mangel, M. 1999. Species–area curves based on geographic range and occupancy. – J. Theor. Biol., 196: 327–342.

Preston, F. W. 1962. The canonical distribution of commonness and rarity: part I. – Ecology 43: 185–215.

Rosindell, J. and Cornell S. J. 2007. Species–area relationships from a spatially explicit neutral model in an infinite landscape. – Ecol. Lett. 10: 586–595.

Rosindell, J. and Cornell, S. J. 2013. Universal scaling of species-abundance distributions across multiple scales. – Oikos 122: 1101–1111.

Simberloff, D. S. and Abele, L. G. 1976. Island biogeography theory and conservation practice. – Science 191: 285–286.

Šizling, A. L. and Storch, D. 2004. Power-law species–area relationships and self-similar species distributions within finite areas. – Ecol. Lett. 7: 60–68.

Šizling, A. L. et al. 2009a. Species abundance distribution results from a spatial analogy of central limit theorem. – Proc. Natl. Acad. Sci. U.S.A. 106: 6691–6695.

Šizling, A. L. et al. 2009b. Invariance in speciesabundance distributions. – Theor. Ecol. 2: 89–103.

Šizling, A. L. et al. 2009c. Rapoport's rule, species tolerances, and the latitudinal diversity gradient: geometric considerations. – Ecology 90: 3575–3586.

Šizling, A. L. et al. 2011. Between Geometry and Biology: The Problem of Universality of the Species–area Relationship. – Am. Nat. 178: 602–611.

Storch, D. et al. 2012. Universal species – area and endemics–area relationships at continental scales. – Nature 488: 78–81.

Tjørve, E. 2010. How to resolve the SLOSS debate: Lessons from species-diversity models. – J. Theor. Biol. 264: 604–612.

Tjørve, E. and Tjørve. K. M. C. 2008. The species–area relationship, self-similarity, and the true meaning of the *z*-value. – Ecology 89: 3528–3533.

Williams, M.R. 1995. An extreme-value function model of the species incidence and species–area relationship. – Ecology 76: 2607–2616.

Willig, M. R. et al. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. – Annu. Rev. Ecol. Syst. 34: 273–309.

Table 1. Notation system.

Notaion	Interpretation
A, R, where $R = \sqrt{A/\pi}$	area of the arena and its virtual radius
a_i	area of the i-th range
$g_{i,j}$, or g_m	area of the j-th gap within the i-th range, or the area of the m-th gap across all ranges
N_g	number of effective gaps
σ, ρ, where $\rho = \sqrt{\sigma/\pi}$	sample plot area and radius
$\text{FAE}(\sigma)$	Finite-Area Effect as a function of scale, i.e., the part of the effective range that would extend beyond the effective arena (Fig. 2c)
$\kappa, \kappa_a, \kappa_g$	shape specific coefficients that relate circumference and area of arena, ranges and gaps
S	species richness
X_e, e.g., A_e	the index 'e' labels the effective values (i.e., areas of effective arena, effective range and effective gap)
X_{tot}	the index 'tot' labels the values that are summed across all species within the arena
X^*	the values rescaled in accord with Storch et al. (2012)
S^{**}	species richness rescaled as suggested here
X', e.g., FAE'	the accent labels the standardized values
$\bar{a}$	mean value
C_t, K_t; where $t = 1, 2, \cdots$	coefficients, that substitute complex expressions

Figure legends:

Fig. 1: Map of an arena (a) and species spatial range with one gap (b). The arena, range and gap are defined by the bold lines. The effective arena is delimited by a thin line that runs around the arena at distance ρ (a). The effective range is the area between the thin full lines that run at the distance of ρ around the range and its gap (b). The bands that make the effective arena and effective gap smaller and effective range larger (the bands between the thin and bold lines) have the width of ρ, and their lengths are c', c'_g and c'_r (dashed lines), respectively. The ρs are the radii of the sample plots (the small circles). The c stands for the circumference of the arena. Black insets show decrease of effective arena (a) and increase of effective range (b) with increasing area of sample plot (small circles). The sample plots on display show (a) that any sample plot with a centre closer to the edge of the arena than ρ would sample the outside of the arena, and (b) that any sample plot outside the range but with its centre closer to the edge of the range than ρ would sample the range.

Fig. 2: The schematic diagrams in four panels show the relationship between effective arena (thin circles) and an effective range (grey areas). The area of the sample plot (σ; $\sigma = \pi \cdot \rho^2$) increases from the panel (a) to the panel (d). Bold lines stand for circular approaches to the arena, range and its gap. The r stands for the radius of the range, which is virtual if the range is not circular (i.e., $r \approx \sqrt{a/\pi}$ where a is an estimation of the 'real' range area). The d stands for the distance between the range and the edge of the arena. The plots a-d refer to Eqs. 12-15, respectively.

Fig 3: Slope of logarithmically-transformed, rescaled SARs ($S^*(\sigma^*)$ as a function of mean area across all ranges of the assemblage in question ($\bar{a}$). Our model predicts that the main driver of variation in slope of $S^*(\sigma^*)$ is $\bar{a}$ (bold hollow curve, Eq. 35). Data (diamonds) were extracted from (a) Lazarina et al. (2013), and (b) Storch et al. (2012) by digitalizing their Figs. 1, and 1 and 2, respectively. Data are more scattered around the predicted relationship, because our derivation ignored effect of scale on the slope of SAR. Full thin lines show linear regressions and the dotted curves their 95% confidence intervals.

Fig. 4: Eleven SARs from Lazarina et al. (2013) and 15 SARs from Storch et al. (2012) (bold symbols) that were rescaled following the Storch et al.'s (2012) (a) and the here introduced (b) rescaling. Only HOL$_\text{I}$ and HOL$_\text{II}$ data (open triangles) as extracted from Lazarina et al. (2013) failed to collapse. Here we used an arithmetic scale for the y-axis and therefore we avoided the problem with decelerated rate of increase in logarithmic functions, which would veil the differences between the data points at large scales.
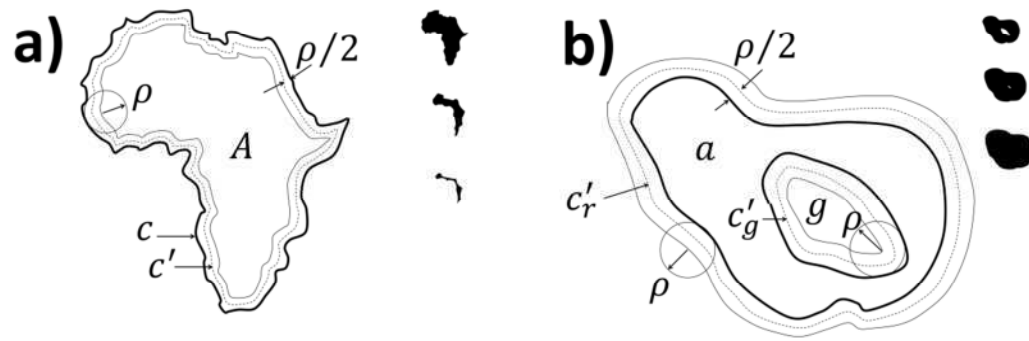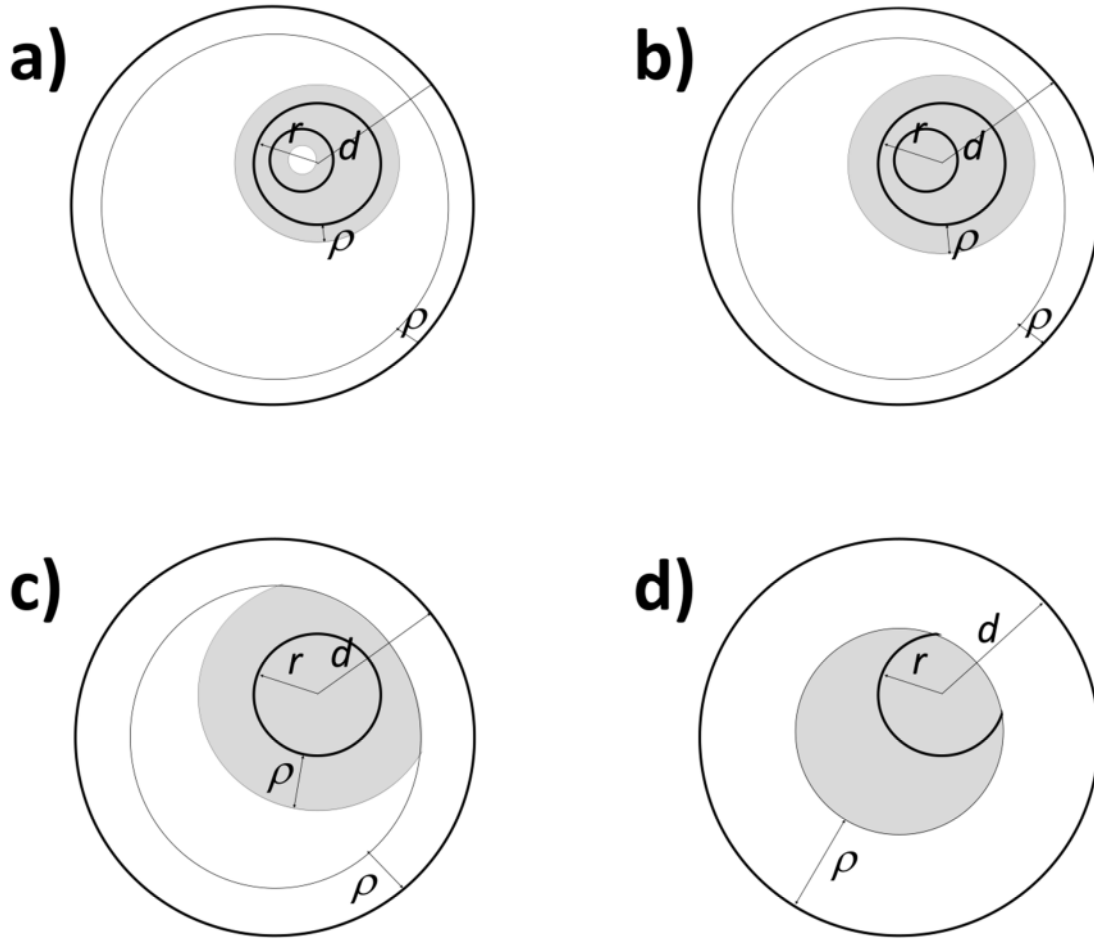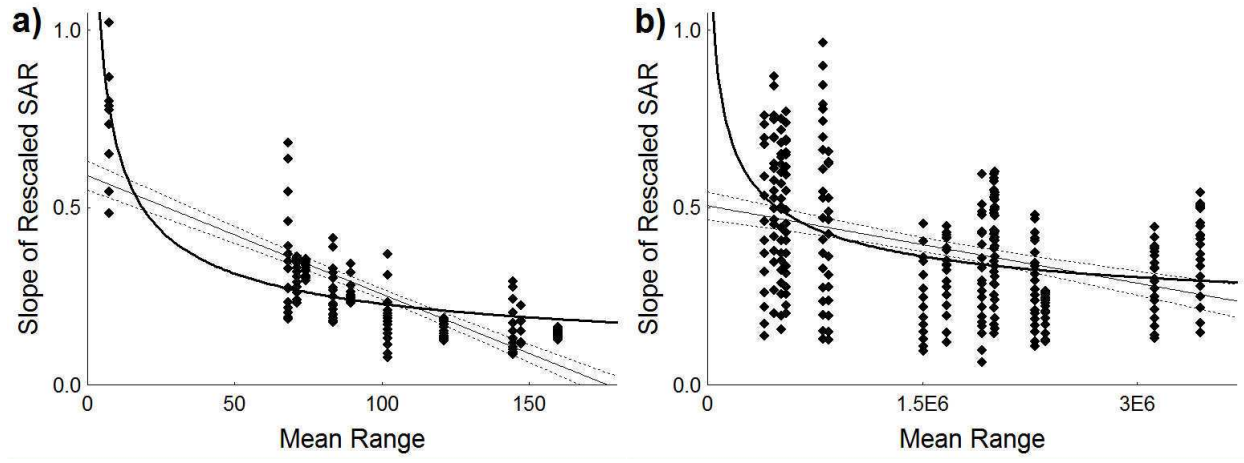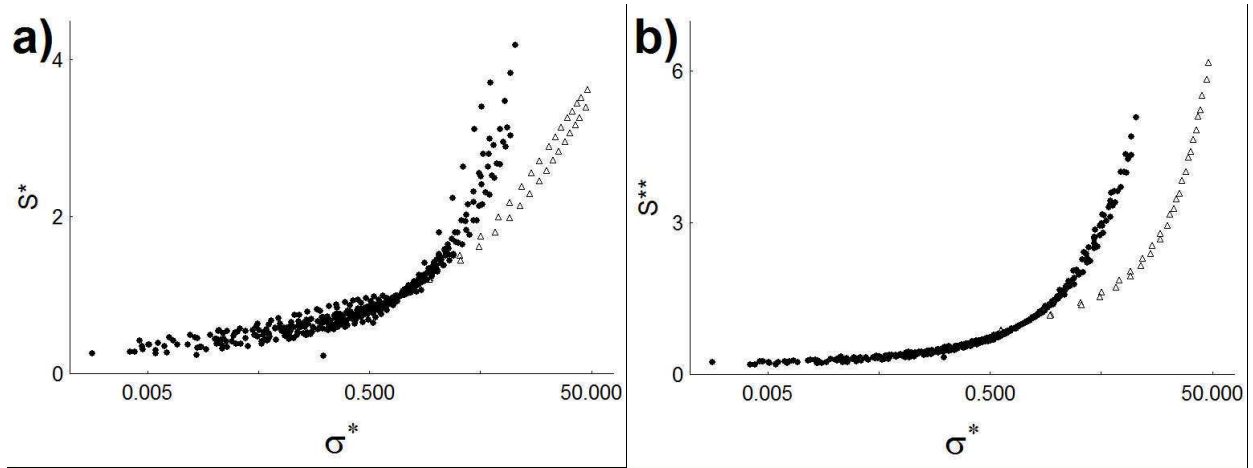
Fig. 1:

Fig. 2:

Fig. 3

Fig. 4