

# Species–area relationship: separating the effects of species abundance and spatial distribution

Even Tjørve<sup>1\*</sup>, William E. Kunin<sup>2</sup>, Chiara Polce<sup>2</sup> and Kathleen M. Calf Tjørve<sup>1</sup>

<sup>1</sup>Lillehammer University College, PO Box 952, N-2604 Lillehammer, Norway; and <sup>2</sup>Faculty of Biological Sciences, Centre for Biodiversity and Conservation, University of Leeds, Leeds LS2 9JT, UK

## Summary

1. The species–area relationship (SAR) has long been used to describe biological diversity across spatial scales. Theoretically, we can describe SARs exhaustively from patterns of species abundance and spatial distribution; where all species are equally common and randomly dispersed, SARs typically rise steeply and then decelerate when plotted on logarithmic axes, but uneven abundances and spatial aggregation lower the SAR curve.
2. In this article, we have developed a generalized model of random placement without replacement for species–occupancy data sets to separate the effects of species abundance and spatial distribution. Observed departures from the model are used to assess the relative contributions of aggregation and unequal occupancy to curve shape. We have applied it both to modelled one- and two-species cases and to plant survey data at both small ( $1 \times 1$  m) and large ( $10 \times 10$  km) spatial resolutions.
3. In modelled distributions, common species raise the SAR most strongly at fine scales, whereas increasingly rare species add to the curve at progressively coarser scales. The effect of aggregation on SAR curvature is greatest with rare species, which lower the curve at most scales, whereas commoner species do so only at fine scales. Uneven occupancies and aggregation may result in convex, concave, S-shaped or inverted S-shaped curves in log–log space.
4. In analyses of field survey data, aggregation contributed less to lowering SAR curvature than did uneven occupancies, both in small and large scale data sets.
5. *Synthesis.* We have thus demonstrated that SAR shape reflects the distribution of species–occupancy and spatial patterning and provide practical methods for parsing the contributions of each to biodiversity across spatial scales. Given the prevalence of gridded occupancy biodiversity data, these methods should have broad application.

**Key-words:** Aggregation, occupancy, random distribution, random-placement model, spatial distribution, spatial pattern, species abundance, species–area relationship, species-diversity,  $z$ -value

## Introduction

The issues of spatial scale and biodiversity have increasingly become topics of ecological interest in recent years (e.g. Tilman & Kareiva 1997; Harte *et al.* 1999; Crawley & Harral 2001; Krishnamari *et al.* 2004; Wilson *et al.* 2004; Drakare *et al.* 2006; Storch *et al.* 2007). However, one of the most important tools for studying these issues, the species–area relationship (SAR; Arrhenius 1921; Gleason 1922), dates back to the foundations of the discipline. Even from its inception, the shape of the SAR has been contentious. The most widely accepted shape is a power–law relationship

(Arrhenius 1921; Rosenzweig 1995), but there has long been both theoretical and empirical support for a range of other shapes (e.g. Gleason 1922; Connor & McCoy 1979; Tjørve 2003). Numerous hypotheses have been presented to account for the shape of species–area curves (see McGuinness 1984; Hill *et al.* 1994; Rosenzweig 1995; Connor & McCoy 2001; Turner & Tjørve 2005 for review), but the issue remains largely unresolved.

In theory, SARs may be described exhaustively from patterns of species abundance and within-species patterns of spatial distribution (aggregation) of individuals. If we have a known number of species ( $S_k$ ) living in a focal area of size  $k$ , the commoner they are on average, and the more evenly they are dispersed within the area, the higher the probability that a

\*Correspondence author. E-mail: even.tjorve@hil.no

small sample within the area will include most or all of them, causing the SAR to rise rapidly towards this number. As species within the area become rarer or more patchily distributed (or both), progressively fewer species are probably found in any subsample, and so the SAR rises less quickly. If there are differences between species in abundance (as there inevitably will be), the slope will vary at different characteristic scales, with commoner species generally being found in small samples, and rare ones in larger ones.

The 'classical' power-law SAR (typically written in the form  $S = cA^z$ , where  $S$  represents species number,  $A$  represents area, and  $c$  and  $z$  are two constants) appears linear when plotted on logarithmic axes ( $\log S$  vs.  $\log A$ ), but an SAR based on random distributions of species (e.g. Coleman 1981; Coleman *et al.* 1982) would deviate markedly from linearity on such axes, rising steeply and then levelling out as it approached  $S_k$ , to form a decelerating curve. Both unequal species abundance and aggregated spatial patterning serve to lower this curve, causing it to rise less steeply and pushing it down towards (or indeed beyond) the power-law SAR. These two mechanistic components (species abundance and spatial patterning) alone should determine the shape of SARs.

Note, however, that this mechanistic approach is not necessarily in competition with other explanations for the shape of SARs based on ecological or evolutionary processes (e.g., habitat diversity, the theory of island biogeography and evolutionary independence; see Turner & Tjørve 2005 for a review). The distinction is between proximate and ultimate causes; indeed the effects of biotic and abiotic factors, evolution, dispersal and extinction on SARs are felt through their effects on species abundances and spatial patterning (He & Legendre 2002).

The general effect of abundance and spatial patterns on curve shape has been documented in previous studies (Solow & Smith 1991; Plotkin *et al.* 2000a; He & Legendre 2002; Green & Ostling 2003; Olszewski 2004; Picard *et al.* 2004; Green & Plotkin 2007). However, the models in these studies are individual-based, examining the effects of species abundance distributions (SADs) and spatial distributions of individuals. Indeed, most methods for the assessment of spatial patterns (by detection of aggregation or clustering) require information on the mapped positions of all individuals within a sample plot (e.g. Condit *et al.* 2000; Plotkin *et al.* 2000a; Tsao 2000; He *et al.* 2002; Picard *et al.* 2004; Green & Ostling 2007) or knowledge of the number of individuals within each quadrat (Solow & Smith 1991). Unfortunately, few distributional data sets are resolved down to the level of individuals and their position in space (e.g. in the tropical-tree census data used by He *et al.* (1997), Condit *et al.* (2000), Plotkin *et al.* (2000a,b) and He & Legendre (2002)); on the other hand, species-occupancy (e.g. gridded distribution) data sets appear with increasing frequency in the literature. Just as species abundance and spatial patterning of individuals determine SAR shape in individual-based data, we should expect species-occupancy and the spatial patterns therein to influence the relationship with gridded distributional data. Indeed, simply by rearranging the pattern of quadrats within

such a species-occupancy matrix, Hopkins (1955) has demonstrated that aggregated occupancy lowers SAR curvature.

#### SPECIES-OCCUPANCY DATA

Species-occupancy data sets typically represent the presence or absence of a species within a rectangular grid of even-sized cells, although in some cases unequal cell sizes may be used (e.g. latitudinal or longitudinal units or geopolitical entities). These may represent occurrences within a limited number of cells (or quadrats) or may cover a substantial geographical area, and records may represent a single intensive survey effort, or they may be accumulated over substantial periods of time. Species-occupancy data sets are widely applied in the study of species diversity, abundance and aggregation. Recently, the relationship between scale and proportional grid occupancy has been explored (Kunin 1998, 2000; Lennon *et al.* 2002; Lennon *et al.* 2007), and grid-occupancy data have been used to study SARs and to discuss self-similarity and power relationships (e.g. Leitner & Rosenzweig 1997; Kunin 1998; Lennon *et al.* 2002; Ulrich & Buszko 2003; Sizing & Storch 2004; Ulrich & Buszko 2004; Ulrich 2005; Lennon *et al.* 2007).

The aggregation of occupied cells in sample grids is analogous to the aggregation of individuals of a species, and the species-occupancy distribution correlates with species abundance (see e.g. He *et al.* 2002; He & Condit 2007 for review), especially when plotted at fine scales of spatial resolution. It has been demonstrated that species-occupancy (or species-occurrence) distributions can follow Poisson and negative binomial distributions (He *et al.* 2002; McGeoch & Gaston 2002; He & Condit 2007). Although the species abundance relationship can be inferred from the occupancy distribution, the actual relationship between the two is not straightforward (see e.g. He *et al.* 2002; Holt *et al.* 2004; He & Reed 2006 for review). McGeoch & Gaston (2002) demonstrated that occupancy-frequency distributions can be either unimodal or bimodal and either relatively symmetrical or skewed. The shape of the occupancy-frequency distribution should be expected to affect the shape of the species-area curve in a predictable manner.

#### THE RANDOM-PLACEMENT MODEL AND OCCUPANCY DATA

A point of departure for studying the role of spatial patterning in SARs can be Coleman's (1981) random-placement model. Coleman *et al.* (1982) reported that the shape of the SAR for breeding birds on islands in Pymatuning Lake indicated random placement (i.e. random spatial distributions). If random placement were the rule in other systems, curve fitting and exploration of abundance would have been simple. Most data sets do not, however, indicate random spatial distributions but rather the aggregation of individuals (see e.g. He *et al.* 1997; Báldi & Kisbenedek 1999; Condit *et al.* 2000). Random distribution may, however, still be regarded as a useful null hypothesis for the spatial pattern in a given data set.

Coleman *et al.*'s (1982) data set provided information about the abundance of species as well as their spatial positions. Measures of spatial pattern and abundance based on grid-occupancy, on the other hand, are independent of the knowledge of species numbers and spatial distribution of individuals. Instead, aggregation in such data is defined as the clustering of occupied grid cells, and abundance is measured as the proportion of cells occupied at a given scale (typically at the finest scale considered).

The aim of this article is to examine systematically the effects and the relative importance of occupancy and spatial patterns on SAR curve shape. If these relationships can be determined successfully, it may also be possible to reverse the process, and use the shape of the SAR to draw conclusions about the relative abundance and spatial patterning of species distributions. This is, however, dependent on an appropriate model being fitted to the data set. In the following, a generalized random-placement model of sampling without replacement is developed for grid-occupancy data (quadrats with occurrence on a sample grid). The effect of relaxing the assumption of random placement and equal occupancy on curve shape is assessed by allowing for the spatial pattern and occupancy distributions found in data sets. This article applies this model to simple, theoretical one- and two-species grid-occupancy relationships to demonstrate the effects of both aggregation and unequal occupancy on curve shape, before using empirical data to test its relevance and utility.

## Methods

### THE MODEL

Individuals of a species are seldom randomly distributed in space (He *et al.* 1997; Condit *et al.* 2000). One way to detect aggregation is to compare actual SARs with SARs from random placement (see e.g. Coleman 1981; Báldi & Kisbenedek 1999; Plotkin *et al.* 2000a; He & Legendre 2002); Coleman (1981), for example, used deviance (more than 2/3 of observations) outside the variance band of the random-placement SAR as an indication of aggregation or regularity (clumped or even distributions).

Because we aim to use grid-occupancy data, Coleman's model in its original form (which depends on knowledge of the number of individuals) is not suitable. Given  $S_k$  is the total number of species in total area  $k$ , and  $a_i$  the number of units of area where species  $i$  occurs, the expected value for the SAR,  $E(S_A)$ , can be alternatively defined as:

$$E(S_A) = S_k - \sum_{i=1}^{S_k} \left(1 - \frac{a_i}{k}\right)^A \quad \text{eqn 1}$$

where  $S_A$  is number of species in area  $A$  (see e.g. Engen 1978). This model is based on sampling with replacement. However, with grid-occupancy data sets, we need to consider sampling without replacement, as the re-sampling of a single cell cannot add species to our accumulated total. A model of sampling with replacement would produce a SAR curve that approaches asymptotically towards the total number of species, whereas a model of sampling without replacement will force the model curve through the given (upper) number of species found in the whole survey area ( $S_k$ ). A hypergeometric distribution expresses such a model of sampling without replacement (see e.g. Ugland *et al.* 2003, eqns 5–7). Again, let  $A$  be

the number of units of area and  $S_k$  the number of species in the total area studied,  $k$ , and  $S_A$  the species in  $A$  units (or grid cells). If  $a_i$  is the number of units where the  $i$ -th species is observed ( $i = 1, \dots, S_A$ ), the expected number of species in a random sample of  $A$  units may be expressed as:

$$E(S_A) = \sum_{i=1}^{S_k} \left[1 - \frac{\binom{k-A}{a_i}}{\binom{k}{a_i}}\right] = S_k - \sum_{i=1}^{S_k} \frac{\binom{k-A}{a_i}}{\binom{k}{a_i}} \quad \text{eqn 2}$$

We may expand this to:

$$E(S_A) = S_k - \sum_{i=1}^{S_k} \left(1 - \frac{a_i}{k}\right) \left(1 - \frac{a_i}{k-1}\right) \dots \left(1 - \frac{a_i}{k-(A-1)}\right) \quad \text{eqn 3}$$

In the special case when all species occupy the same amount of area, i.e. all the occupancies  $a_i$  are equal to a common occupancy  $a$ , these expressions simplify to

$$E(S_A) = S_k \left[1 - \frac{\binom{k-A}{a}}{\binom{k}{a}}\right] \quad \text{eqn 4}$$

and

$$E(S_A) = S_k \left[1 - \left(1 - \frac{a}{k}\right) \left(1 - \frac{a}{k-1}\right) \dots \left(1 - \frac{a}{k-(A-1)}\right)\right] \quad \text{eqn 5}$$

The difference between eqn 3 and eqn 5 then constitutes the effect of differences in species-occupancy on the SAR but still assumes a random spatial distribution. We can obtain the (expected) number of species with an even occupancy, but with the observed rather than random spatial distribution, by adding the difference between models 3 and 5 to the observed number of species,  $S$ :

$$E(S_A) = S + \sum_{i=1}^{S_k} \left(1 - \frac{a_i}{k}\right) \left(1 - \frac{a_i}{k-1}\right) \dots \left(1 - \frac{a_i}{k-(A-1)}\right) - S_k \left(1 - \frac{a}{k}\right) \left(1 - \frac{a}{k-1}\right) \dots \left(1 - \frac{a}{k-(A-1)}\right) \quad \text{eqn 6}$$

Both a random spatial distribution of individuals (assuming zero-size individuals) and a random pattern of coverage (with space-filling individuals) of species will result in a random spatial-occupancy pattern within a grid set. With random occupancy, the difference between models 3 and 5 can only be a result of a difference in the distribution of occupancies. The two models, eqns 3 and 5, thus serve to separate out the effect of non-random spatial distributions and of an unequal species-occupancy distribution on curve shape.

### ARTIFICIAL DATA

To demonstrate the effect of aggregation and species-occupancies (represented by occupancy distributions) on SAR shape, we first apply model 3 to a simple, artificial grid-occupancy relationship generated from a 6 × 6-cell grid, with one or two species with different degrees of aggregation and grid occupancy. The number of quadrats occupied at each scale is calculated from all continuous possibilities including wrapped-around sampling areas (as done by Leitner & Rosenzweig 1997).

### THE FIELD DATA

We next apply the model to field-collected data, using plant survey data collected at two very different spatial scales: small-scale

**Table 1.** Descriptions of the survey data sets considered, including location (on the British Isles), and altitude. The small-scale data set has  $1 \times 1$  m cells, whereas the large-scale data set has  $10 \times 10$  km ( $100 \text{ km}^2$ ) cells. The table also shows the total number of species ( $S_k$ ) and the mean number of occupied cells (across all species)

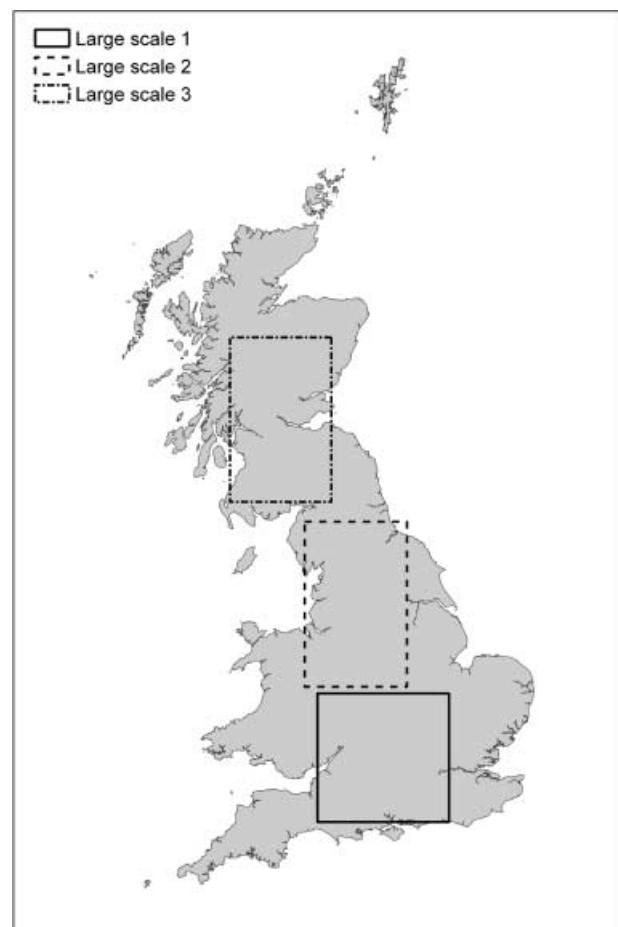
Data set	Description	Localization	Altitude	$S_k$	mean
Small scale 1	Heathland	New Forest	(low altitude)	34	211.2
2	Mountain grassland	Lake District	1300ft	54	135.4
3	Beechwood	Chiltern Hills	(low altitude)	28	59.6
4	Blanket bog	near Westport	(low altitude)	48	162.0
5	<i>Calluna</i> bog	Perthshire	950ft	47	164.1
6	Pine forest	Perthshire	900ft	40	99.9
7	<i>Zostera</i> bog	Norfolk	low tide level	3	155.3
8	<i>Salicornia</i> marsh	Norfolk	high-tide level	5	179.6
9	Shingle ridge (beach)	Norfolk	above tide level	18	94.3
10	<i>Calluna</i> moor	Church Stretton	1450ft	28	186.1
11	<i>Calluna</i> bog	North Pennines	1800ft	46	145.1
12	Limonium marsh	Norfolk	well above high tide	14	288.7
Large scale 1	na	Southern England	na	1863	139.6
2	na	Northern England	na	1835	108.0
3	na	Scotland	na	1665	87.7

( $1 \times 1$  m resolution) and large-scale ( $10 \times 10$  km resolution) surveys. (Note that we will use the words ‘small’ and ‘large’ to distinguish these two data set scales, and the words ‘fine’ and ‘coarse’ to distinguish different scales of analysis of a given data set).

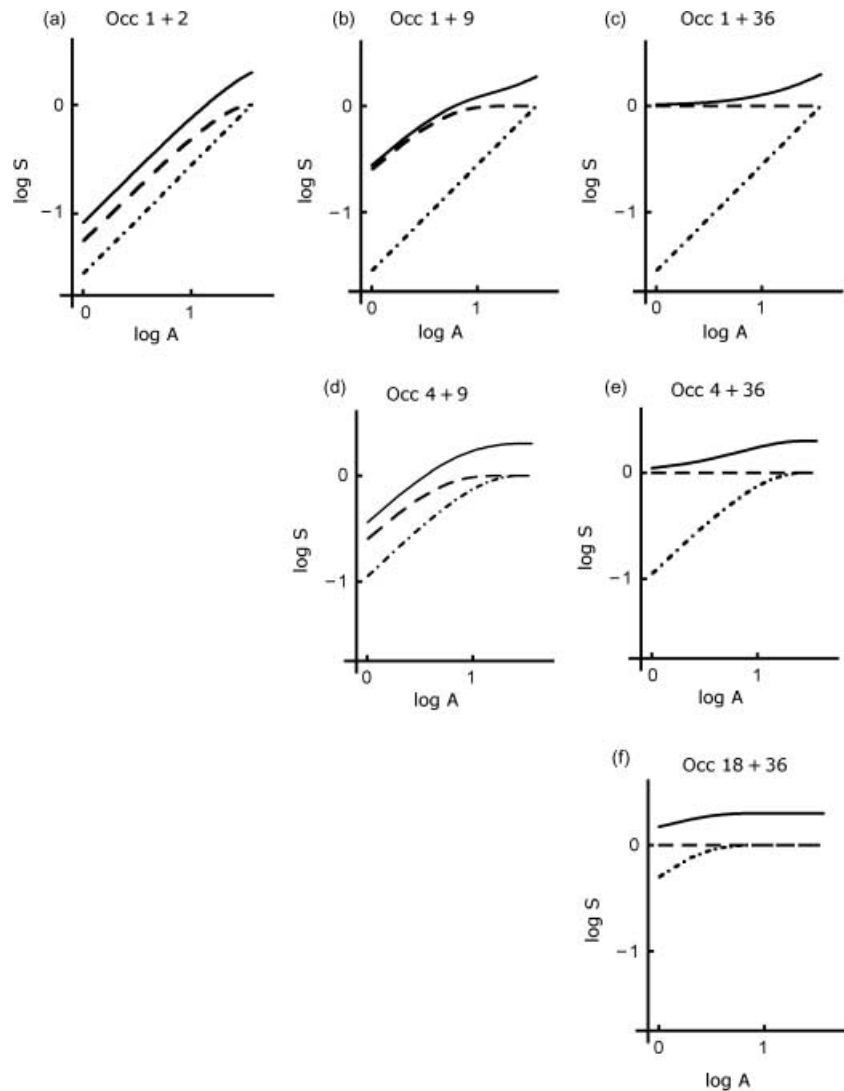
There has been considerable interest recently in the scale-dependence of ecological phenomena (e.g. Shmida & Wilson 1985; Kunin 1997; Storch *et al.* 2008); by considering two parallel sets of analyses differing by eight orders of magnitude in area, we hoped to test for the generality or of the patterns revealed. Hopkins (1955) collected the small-scale data to examine SARs, and the data were subsequently re-analyzed (Kunin 1997) in a study of sample-shape effects. We have taken the large-scale data from the recently published *Millennium Atlas of the British Flora* (Preston *et al.* 2002). Hopkins’ (1955) patch-occupancy data set consists of occurrence data from  $20 \times 20$  m matrices of  $1 \times 1$  m squares from 12 semi-natural plant communities in the British Isles (Table 1). Species-occurrences were recorded for all 400  $1 \times 1$  m squares in each matrix, and covered all vascular plants, bryophytes and lichens. To make a comparable large scale data set, we have chosen three rectangular areas of data from the British Atlas, each comprising 400 cells (Fig. 1). Due to the shape of the land mass, only the southernmost of these (large scale 1) was square ( $20 \times 20$  cells) with the other two somewhat elongated ( $25 \times 16$  cells; large scale 2 and 3 from south to north).

From each of these 400 cell arenas, we sampled species-richness at 8 scales:  $A = 1, 4, 16, 25, 64, 100, 256$ , and 400 cells ( $n = 400, 100, 25, 16, 8, 4, 4$  and 1, respectively). Non-overlapping square samples were taken at each scale, except at the coarsest scales ( $A = 256$ , and in elongated sampling arenas also  $A = 100$ ), where the minimum possible overlap was employed.

In each case, we used these grid-occupancy data to produce hypothetical SARs for (i) random distribution and equal occupancy (model 5) and (ii) observed occupancies with random distributions (model 3) to assess the effect of spatial patterns and unequal occupancy on curve shape. Using the equations above, we calculated data points for observed and modelled species-diversity. Proportional contributions from unevenness in occupancy and aggregation were calculated as proportions of model-5 values. We plotted the number of cells occupied for each species in increasing order. GraphPad Prism 4 and Agrapher 2.11 software were employed for plotting the curves.



**Fig. 1.** Locations of the three large-scale data sets (on the British Isles). The large-scale data set is made up of sets of  $10 \times 10$  km cells. Each set includes 400 cells, although the shapes are varied somewhat due to the shape of the land mass.



**Fig. 2.** (a–f) SARs in log–log space from pairs of randomly distributed species in universes of 36 cells. Occupancies for the two species in each figure (a–f) are given above each figure. The dashed-dotted line represents the rarest species, the dashed line the commonest species, and the unbroken line the resulting SAR.

We did not fit the power curve and the exponential curve by regression; instead, for the sake of illustrative comparison, we drew them between the points for the mean  $S$  value of a single quadrat and for that of the total sample grid, that is, between the values for  $A = 1$  as the starting point and  $A = 400$  as the end point. Then, for such a grid-occupancy data set (from a sample grid), the  $c$ -value of the power-law SAR is determined by the mean number of species in one cell, and the slope in log–log space is calculated as  $z = (\log S_k - \log c) / \log k$ . Therefore, given a fixed number of total species, a higher  $c$ -value reduces the  $z$ -value and vice versa, and the average proportion of species found in one cell is given by  $c/S_k$ .

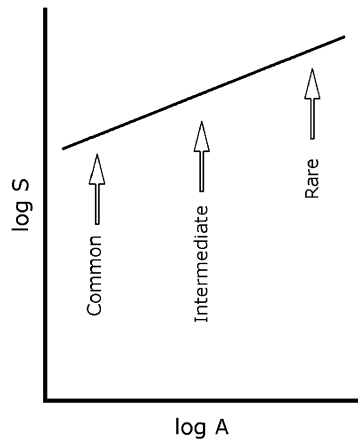
## Results

### ARTIFICIAL DATA

Varying occupancies in two-species model universes (of  $6 \times 6$  cells) illustrate how abundance affects curve shape. In Fig. 2a–f, occurrences are varied systematically for two species, and the resulting SARs plotted.

Each species lifts part of the curve according to its abundance. Rare species push up the SAR the most at the right-hand (high  $A$ ) side, generally increasing SAR slope, whereas common species lift the whole curve, generally lowering the overall slope. Combining species of similar occupancy (2a, d or f) always produces a decelerating SAR, but combining species of dissimilar occupancy (e.g. 2c, but also 2b or 2e) can produce accelerating curvature at some scales. Taken together, these plots indicate how even a very simple assemblage of two species can display a wide range of SAR shapes, including both sigmoid and inverted sigmoid curves, depending on the relative levels of occupancy of the species involved (see also Fig. 3).

Spatial aggregation in occupancy lowers the SAR at intermediate scales, while leaving the two ends (single-cell and full-grid species counts) unaffected. It can thus move the decelerating SAR of random placement down towards (or even beyond) the diagonal of the power law. As the SAR is the sum of multiple species–area curves (Kunin 1998), the effect of aggregation can be demonstrated using such plots.



**Fig. 3.** The effect of abundance (or occupancy) on curve shape. The arrows indicate where common, rare, and intermediate species will bend the curve while lifting it upward.

Aggregation has the greatest effect when it occurs in rare species, lowering the curve across a wide range of scales (Fig. 4a), whereas for commoner species the curve is affected only at relatively fine scales (Fig. 4b–c). The SAR of a multi-species assemblage (i.e. the sum of such curves) should thus be most sensitive to spatial aggregation of the rarest

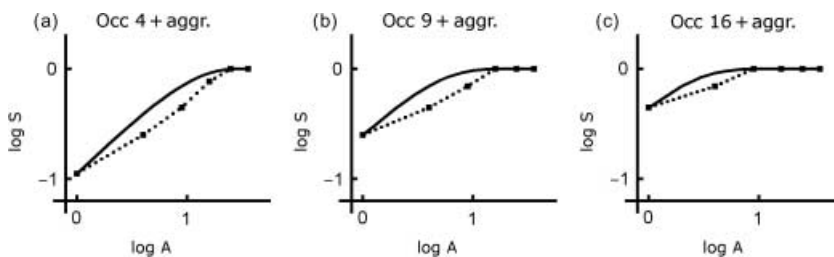
species in the assemblage. Aggregated distributions in these cases contribute towards a (more) sigmoid rather than a decelerating (convex-upward) curve.

These principles above should apply to any SAR, making it possible to interpret abundance or occurrence distribution based on its curve shape and its starting point (which indicates average species density in one cell). The presence of many rare or many common species is also revealed by the mean occupancy at the finest scale, in addition to the course of the curve.

#### FIELD SURVEY DATA

In the small-scale field data sets (plant surveys), the total number of species varied from 3 to 56 at each site, and the mean number of occupied cells (across all species) varied between 60 and 289 (of 400). In the large-scale field data, the total species number ranged from 1665 to 1863, and mean occupancy ranged from 88 to 140 (see Table 1).

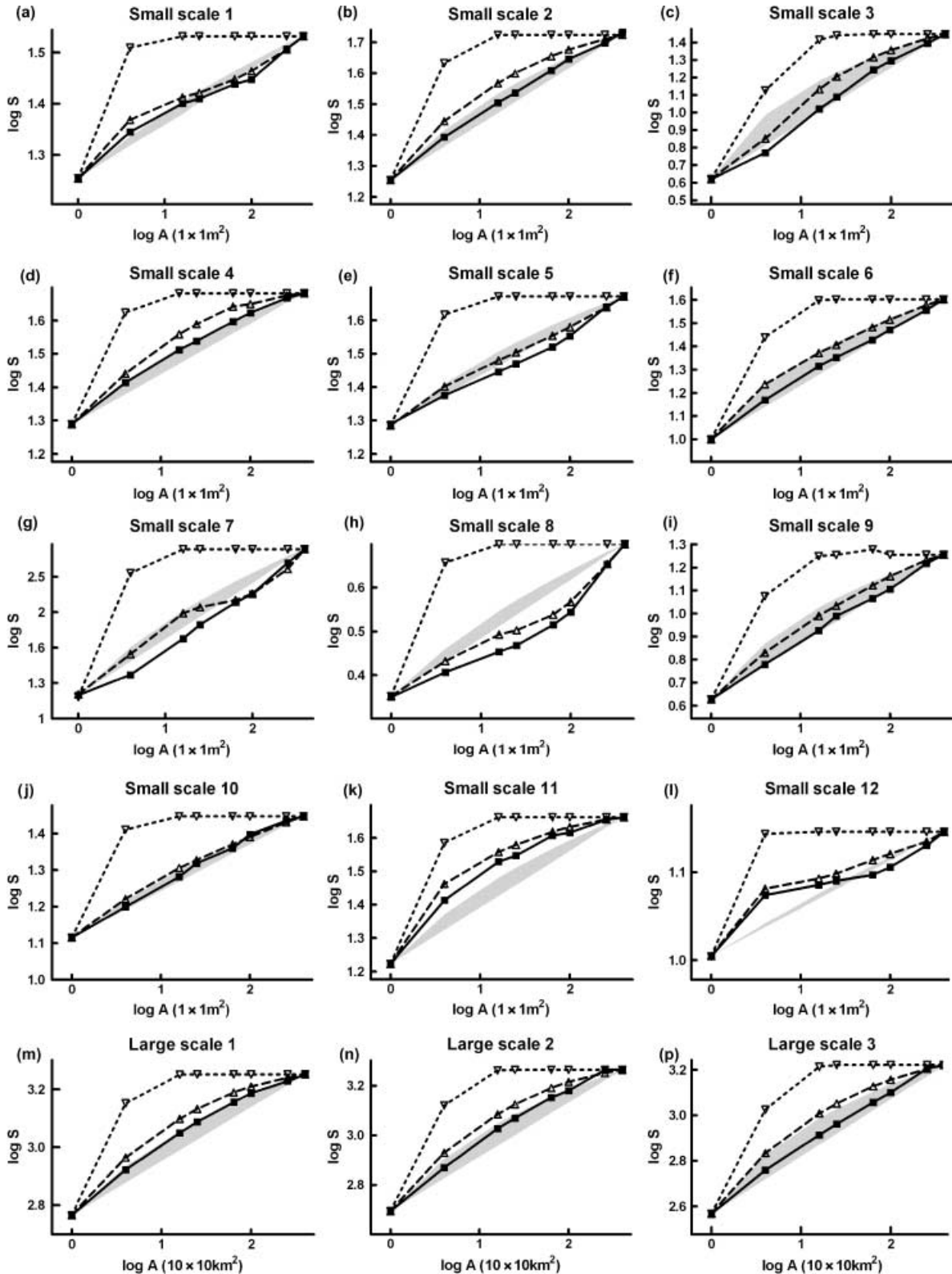
The log–log SAR plots for all small-scale and large-scale surveys are found in Fig. 5a–l and 5m–p, respectively. The graphs show the observed curve compared to the model curve for random spatial distribution but with observed levels of occupancy (model 3), and the model curve for random distribution and equal occupancies for all species (model 5).



**Fig. 4.** (a–c) Illustrates how aggregation affects curve shape for a single species. The unbroken curve represents random distribution (model 5), and the dotted curve represents maximum aggregation (i.e. all cells with occurrences aggregated into a randomly positioned block).

**Table 2.** Proportional lowering of the SAR curves from unevenness in occupancy (Occu.) and aggregation (Aggr.), for  $A = 4, 16, 25, 64, 100$  and 256. More than twofold differences are indicated by bold-faced type

Site	$A = 4$		$A = 16$		$A = 25$		$A = 64$		$A = 100$		$A = 256$	
	Occu.	Aggr.	Occu.	Aggr.	Occu.	Aggr.	Occu.	Aggr.	Occu.	Aggr.	Occu.	Aggr.
Small scale 1	<b>0.28</b>	0.04	<b>0.24</b>	0.02	<b>0.22</b>	0.02	<b>0.17</b>	0.02	<b>0.14</b>	0.03	<b>0.06</b>	0.00
2	<b>0.35</b>	0.07	<b>0.30</b>	0.09	<b>0.25</b>	0.10	0.15	0.09	0.11	0.06	0.03	0.03
3	<b>0.47</b>	0.09	<b>0.48</b>	0.12	<b>0.42</b>	0.14	<b>0.26</b>	0.11	0.19	0.10	0.05	0.06
4	<b>0.34</b>	0.04	<b>0.25</b>	0.08	<b>0.19</b>	0.09	0.09	0.09	0.07	0.05	0.02	0.02
5	<b>0.39</b>	0.04	<b>0.36</b>	0.05	<b>0.32</b>	0.05	<b>0.24</b>	0.06	<b>0.19</b>	0.05	<b>0.07</b>	0.00
6	<b>0.37</b>	0.09	<b>0.41</b>	0.07	<b>0.36</b>	0.07	<b>0.24</b>	0.09	<b>0.18</b>	0.08	0.06	0.04
7	<b>0.41</b>	0.07	<b>0.34</b>	0.10	<b>0.31</b>	0.07	<b>0.28</b>	0.01	<b>0.25</b>	0.00	<b>0.12</b>	–0.04
8	<b>0.41</b>	0.03	<b>0.38</b>	0.05	<b>0.36</b>	0.05	<b>0.31</b>	0.04	<b>0.26</b>	0.04	<b>0.10</b>	0.00
9	<b>0.43</b>	0.06	<b>0.45</b>	0.07	<b>0.40</b>	0.06	<b>0.30</b>	0.08	0.19	0.10	0.05	0.03
10	<b>0.35</b>	0.03	<b>0.28</b>	0.04	<b>0.24</b>	0.02	<b>0.16</b>	0.02	<b>0.12</b>	–0.02	<b>0.04</b>	–0.01
11	<b>0.25</b>	0.08	<b>0.21</b>	0.05	<b>0.18</b>	0.06	<b>0.10</b>	0.02	0.07	0.04	0.01	0.00
12	<b>0.13</b>	0.01	<b>0.12</b>	0.01	<b>0.10</b>	0.02	0.07	0.04	0.06	0.03	<b>0.03</b>	0.01
Large scale 1	<b>0.35</b>	0.06	<b>0.30</b>	0.07	<b>0.24</b>	0.07	0.13	0.07	0.09	0.05	0.03	0.03
2	<b>0.35</b>	0.08	<b>0.33</b>	0.08	<b>0.28</b>	0.09	<b>0.15</b>	0.07	0.11	0.07	<b>0.03</b>	–0.03
3	<b>0.36</b>	0.10	<b>0.38</b>	0.12	<b>0.32</b>	0.13	0.20	0.12	0.14	0.11	<b>0.04</b>	–0.01



**Fig. 5.** (a–p) SARs for small-scale and large-scale empirical data sets presented on log–log axes. Fig. 5a–l show SARs from Hopkins' (1955) 12 (small-scale) sites, and Fig. 5 m–p show SARs from 3 large-scale ( $200 \times 200$  km/ $125 \times 80$  km) matrices taken from the *Millennium Atlas of the British Flora* (Preston *et al.* 2002). The grey shade represents the difference between the power model (straight line) and the exponential model (convex upward curve) drawn between the average number of species in single sample quadrat and the total number of species in the full 400 cell arena. The solid line and filled squares represents the observed SAR with calculated scales ( $A = 1, 4, 16, 25, 64, 100, 256$  and 400) indicated as squares. The dashed line with upward facing triangles represents model 3 with random spatial distribution and the dotted line with downward triangles represents model 5 with random spatial distribution and equal occupancy. Thus the area between the solid and dashed lines represents the effect of aggregation, whereas the area between the dashed and dotted lines represents the effect of uneven occupancy.

It should be noted that small-scale sites 7 and 8 contain very few species (3 and 5 species, respectively), and their SARs may, therefore, be less definitive than those of the other sites. Proportional lowering of the curve (relative to the model-5 curve) caused by unevenness in occupancy and aggregation of occurrence is given in Table 2.

The 12 small-scale survey SARs exhibit different shapes and deviations from the power model, as do those of the 3 large-scale data sets. For some sites, e.g. small-scale site 6 and 9 and large-scale site 3, the power relationship fits quite well whereas for small-scale site 4 and large-scale site 1 the exponential relationship seems to produce a better fit, and for small scale site 2 the curve falls between the two models. However, most sites (especially at the small scale) display curves outside the area covered between the power model and the exponential model.

Model-3 curves are higher (generally more decelerating) than the observed curve, indicating the effect of aggregated spatial distributions on curve shape. Model-5 curves are even higher and are always sharply decelerating in shape, as the effect of uneven occupancy is also removed. The two models illustrate how both abundance and spatial pattern (in this case aggregation) lower and reshape the SAR. In all sites and at both scales (small-scale and large-scale data sets), occupancy appears to be more important than aggregation in explaining the observed shape: unequal occupancy lowers the curve more in all (but perhaps the very coarsest) scales considered. In sites 5 and 8, unequal occupancy by itself is sufficient to lower the SAR curve down to or below the power relationship.

The model-3 curves have inverted sigmoid shapes for several of the sites, but the model-5 curves never have. This indicates that uneven occupancy rather than aggregation is the main contributor to such S-shapes in SARs.

Low curvatures in model-3 curves, as found in small-scale sites 5 and 8, indicate more uneven occupancies, that is, a large proportion of common and rare species; the high SAR curvatures, as in small-scale sites 4 and 11, indicate more even levels of occupancy. The inverted S-shapes compressed towards the finer-scale (lower) end of the curve, as in small-scale sites 1 and 12, indicate many common species; the inverted S-shape compressed towards the coarser-scale end, as in small-scale site 3, indicates many rare species.

#### OCCUPANCY DISTRIBUTIONS

We may compare the predictions above (made from curve shapes) with the actual occupancy distributions found in the grid data sets. The Supporting Information provides the number of cells occupied plotted for each species (see Appendix S1 in Supporting Information). These plots also clearly illustrate how abundance affects curve shape. For example, small-scale data sets 1 and 12 show a high proportion of species with relatively high occupancies and few at medium. This effect is visible in the plots (in Fig. 5a and l) as a pronounced decelerating shape at low to medium scales and an inflection point at medium scales (as indicated above), but the curve at coarser scales accelerates owing to the influence of

some species with very low occupancies, thus producing an inverted S-shaped SAR. According to the model data, we can expect ordinary S-shapes if there are few fairly common to very common species but many rare to relatively rare species. These conditions are found in small-scale data set 3. Small-scale site 8 has very few species, but most of them are either very common or very rare, resulting in a distinctly accelerating SAR. Small-scale site 11, on the other hand, has mostly species with medium occupancies that predictably cause the curve to become distinctly decelerating. The large-scale data sets (as illustrated in the appendix) tend to have fewer medium-common species towards the north, causing the curvature of the SAR to become flatter (Fig. 5m–p). Whereas the curve for the southernmost grid is close to the exponential model, the curve for the northernmost is much closer to a power relationship.

According to the modelled data (Fig. 4), the presence of rare species mainly causes the lowering of the curvature by aggregation. Aggregation in commoner species only affects the curve at relatively fine scales. This is evident in the aggregations affecting curve shape at most scales in small-scale data sets 3, 6 and 9, which all have large proportions of rare species, but the trends are not as clear as those caused by differences in occupancy patterns.

#### Discussion

Kunin (1998) has presented a method (also applied by Kunin 2000; Lennon *et al.* 2002; Lennon *et al.* 2007) that gives a cross-scale measure of a single species abundance visualized as scale–area plots in log–log space (also termed ‘Range–area relationships’ (Ostling *et al.* 2000) or ‘P-area curves’ (Storch *et al.* 2008)). Both the species abundance or occupancy (i.e. the number of cells occupied at the finest scale considered) and the spatial aggregation or dispersion of that occupancy determine this relationship between the area deemed to be occupied by a species and the spatial scale of analysis. The single-species curves shown in Figs 2 and 4 are effectively scale–area plots, which can be interpreted as showing the probabilities that a subplot of a given area will contain the species in question. The sum of such probabilities across species determines the expected number of species to be found in a given area,  $S_A$ , and thus provides us with the SAR. Just as occupancy and spatial patterning determine single-species curves, these two variables as expressed in the entire species set determine this summed curve. Several authors have found that uneven SADs lower the curvature of this summed SAR (Solow & Smith 1991; Plotkin *et al.* 2000a; He & Legendre 2002; Green & Ostling 2003; Olszewski 2004; Picard *et al.* 2004). We have extended this work, showing both how it can be applied to gridded presence–absence data and how the relative contributions of unequal occupancy and spatial patterning to SAR shape can be distinguished.

We applied the model (and equations) presented here to two very different groups of British plant data sets (at 1 m and 10 km linear resolutions), and the results consistently showed that occupancy (abundance) distribution contributed more



to a lowered the SAR curve than did spatial aggregation (clumping of individuals) (Table 2). We thus conclude that, after the number of species and overall density (the sum of occupancies across all species), variation in abundance or occupancy contributed most to explaining curve shape. The effect of abundance on curve shape varies both with the number of common and rare species. Both model and empirical data sets illustrate well how unequal occupancy in grid-occupancy SARs tends to move the curve from the steeply decelerating shape implied by random placement towards the linear power-law, or even beyond it to form an accelerating curve. Skewed occupancies may produce both S-shaped and inverted S-shaped curves. Aggregation may also contribute to S-shapes, particularly when there are many rare species. The importance of rare species in aggregation effects (see Fig. 4) is logical; when using occupancy data, common species with very large coverage cannot display much clumping beyond the finest scales of analysis, as their gridded distributions quickly become 'saturated' at coarser scales. It is not clear whether similar trends appear when different (individual-based) aggregation indices are employed; Condit *et al.* (2000) have found that rare species may be more aggregated than common species, whereas Plotkin *et al.* (2000a) have found no correlation between aggregation and abundance.

The findings from our simply modelled distributions coincide well with the results from the empirical data sets. The fact that  $c/S_k$  correlates well with proportion of species found in more than 350 cells ( $R^2 = 0.67$ ,  $P = 0.0002$ ) illustrates the effect of many common species on  $c$  (of the power model). A comparison between the observed curves (Fig. 5a–p) also reveals that data sets with a proportion of species found in more than 350 cells (or high  $c/S_k$ ) consistently have the most decelerating SARs.

#### THE ISSUE OF GENERALITY

As noted above, one of the most striking results reported here is that unequal species-occupancy was found to have a much stronger effect on SAR shape than did the spatial patterning of those distributions. This proved to be the case both in very fine-scale (1 m<sup>2</sup> resolution) and very coarse-scale (100 km<sup>2</sup> resolution) data sets, suggesting that this property is relatively invariant across scales. Nonetheless, it is unclear whether the primacy of unequal occurrence over aggregation is a general rule or a specific property of the system examined (British plants). One might find that in more species-rich (e.g. tropical) environments or with more mobile organisms (e.g. butterflies), rather different relative weightings might occur. In addition, it should be noted that the ratio of data grain (resolution) and data-set extent was the same for both sets of analyses (400 cells); it could be that spatial-aggregation effects might be stronger when tested in larger arenas. Interestingly, we found that the proportional lowering of the curvature from unevenness in occupancy and aggregation was consistent between the large- and small-scale data sets. Still, it might be worth investigating whether this consistency holds also for different ratios of grain to extent.

At coarse and intermediate scales, occupancy in a grid should approximate effectively patterns of aggregation in nature caused by spatial distribution of individuals. As cell size becomes very small, however, they may approach or even become smaller than the sizes of single individuals. If individuals are recorded as points, one should anticipate encountering some type of evenness (in addition to aggregation), as each individual in reality occupies space. If occupancy is defined as any part of an individual being inside the cell, the expected SAR will approach asymptotically towards the sum of proportional cover across all species. Either way, the SAR can be expected to behave in 'strange ways' at very small scales.

#### DEVIATIONS FROM THE POWER MODEL

Because SADs and spatial patterns differ between communities (May 1975), the discovery of a universally best regression model for SARs is most unlikely. The non-linearity observed in log–log space, including sigmoid shapes, indicates that additional models other than the power and the exponential models might produce a better fit at the scales studied (see Tjørve 2003 for review of possible models). The present data set seems to indicate that, if the power model is still useful as a first approach, the different data sets equally produce curves on both sides of the power relationship. The power model, therefore, approximates the SARs over limited ranges, as He & Legendre (2002) have also observed. This notion can be stated as the z-value varying with area (He *et al.* 1996; Crawley & Harral 2001; Fridley *et al.* 2005; He *et al.* 2005).

Even though we found that uneven occupancies between species depress the SAR curve about five times more than aggregation in our plant grid-occupancy data sets, aggregation is still important in shaping the curve. Consequently, searches for the SAD that best fits the power SAR are not likely to provide much predictive power, as such approaches typically do not incorporate spatial aggregation into the analyses.

#### Conclusion

Both Crawley (1997) and He & Legendre (2002) have argued both that species abundance and spatial distribution are the two most important factors in interpreting species-diversity and that we need to improve our understanding of their relative influences. The generalized random-placement model with sampling without replacement, as presented in this article, provides a useful step towards disentangling these two effects. It may be applied whenever fractional occupancy is studied in gridded sample areas, and it assesses the relative contribution of these two factors to SAR shape.

Model 5 provides an idealized SAR for random distribution and equal occupancy. Whereas model 3 removes the effect of uneven occupancy alone, model 6 removes the effect of aggregation. Plotting (two or more of) the three functions thus allows the parsing of SARs into these two key component processes, making it possible to calculate the relative importance of spatial patterns and abundance in sample-grid data sets.

For both the large-scale and the small-scale data sets, aggregation contributed less to lowering SAR curvature than did uneven occupancies. Still, differences in occupancies and aggregation may both contribute to or bring about different shapes: convex, concave, S-shaped or inverted S-shaped curves in log–log space. Consequently, aspects of biodiversity patterning, such as abundance and aggregation, should be predictable from the SAR shape, suggesting fruitful directions for future biodiversity research.

## Acknowledgements

Authors thank Karl Inne Ugland and Sigmund Andreassen for invaluable help with the models, Jorge Ramirez for help with programming (automated routines), and Steven Connolly for editing and correcting the language.

## References

- Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.
- Báldi, A. & Kisbenedek, T. (1999) Orthopterans in small steppe patches: an investigation for the best-fit model of the species-area curve and evidences for their non-random distribution in the patches. *Acta Oecologica*, **20**, 125–132.
- Coleman, B. (1981) On random placement and species-area relationships. *Mathematical Biosciences*, **54**, 191–215.
- Coleman, B.D., Mares, M.A., Willig, M.R. & Hsieh, Y.-H. (1982) Randomness, area and species richness. *Ecology*, **64**, 1121–1133.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N. *et al.* (2000) Spatial patterns in the distribution of tropical forest trees. *Science*, **288**, 1414–1418.
- Connor, E.F. & McCoy, E.D. (1979) The statistics and biology of the species-area relationship. *American Naturalist*, **113**, 791–833.
- Connor, E.F. & McCoy, E.D. (2001) Species-area relationships. *Encyclopedia of Biodiversity*, Vol. 5 (ed. S.A. Levin), pp. 397–411. Academic Press, San Diego.
- Crawley, M.J. (1997) The structure of plant communities. *Plant Ecology*, 2nd edn. (ed. M.J. Crawley), pp. 475–531. Blackwell, Oxford.
- Crawley, M.J. & Harral, J.E. (2001) Scale dependence in plant biodiversity. *Science*, **291**, 864–868.
- Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of geographical, evolutionary and ecological context on species-area relationships. *Ecology letters*, **9**, 215–227.
- Engen, S. (1978) *Stochastic Abundance Models*. Chapman and Hall, London.
- Fridley, J.D., Peet, R.K., Wentworth, T.R. & White, P.S. (2005) Connection fine- and broad-scale species-area relationships of southeastern U.S. flora. *Ecology*, **86**, 1172–1177.
- Gleason, H.A. (1922) On the relation between species and area. *Ecology*, **3**, 158–162.
- Green, J. & Ostling, A. (2007) Endemics-area relationships: the influence of species dominance and spatial aggregation. *Ecology*, **84**, 3090–3097.
- Green, J.L. & Ostling, A. (2003) Endemics-area relationships: the influence of species dominance and spatial aggregation. *Ecology*, **84**, 3090–3097.
- Green, J.L. & Plotkin, J.B. (2007) A statistical theory for sampling species abundances. *Ecology letters*, **10**, 1037–1045.
- Harte, J., McCarthy, S., Taylor, K., Kinzig, A. & Fisher, M.L. (1999) Estimating species-area relationships from plot to landscape scale using species spatial-turnover data. *Oikos*, **86**, 45–54.
- He, F. & Condit, R. (2007) The distribution of species: occupancy, scale, and rarity. *Scaling Biodiversity* (eds D. Storch, P. Marquet & J.H. Brown), pp. 32–50. Cambridge University Press, Cambridge.
- He, F., Gaston, K.J. & Wu, J. (2002) On species occupancy-abundance models. *Ecoscience*, **9**, 119–126.
- He, F. & Legendre, P. (2002) Species diversity patterns derived from species-area models. *Ecology*, **83**, 1185–1198.
- He, F. & Reed, W. (2006) Downscaling abundance from the distribution of species. *Scaling and Uncertainty Analysis in Ecology: Methods and Applications* (eds J. Wu, K.B. Jones, H. Li & O.L. Loucks), pp. 89–108. Springer, Dordrecht, The Netherlands.
- He, F., Gaston, K.J., Connor, E.F. & Srivastava, D.S. (2005) The local-regional relationship: immigration, extinction, and scale. *Ecology*, **86**, 360–365.
- He, F., Legendre, P. & Frankie, J.V. (1997) Distribution patterns of three species in a Malaysian tropical rain forest. *Journal of Vegetation Science*, **8**, 105–114.
- He, F., Legendre, P. & LaFrankie, V. (1996) Spatial patterns of diversity in a tropical rain forest of Malaysia. *Journal of Biogeography*, **23**, 57–74.
- Hill, J.L., Curran, P.J. & Foody, G.M. (1994) The effect of sampling on the species-area curve. *Global Ecology and Biogeography Letters*, **4**, 97–106.
- Holt, A.R., Warren, P.H. & Gaston, K.J. (2004) The importance of habitat heterogeneity, biotic interactions and dispersal in abundance-occupancy relationships. *Journal of Animal Ecology*, **73**, 841–851.
- Hopkins, B. (1955) The species-area relations of plant communities. *Journal of Ecology*, **43**, 409–426.
- Krishnamari, R., Kumar, A. & Harte, J. (2004) Estimating species richness at large spatial scales using data from small discrete plots. *Ecography*, **27**, 637–642.
- Kunin, W.E. (1997) Sample shape, spatial scale and species counts: implications for reserve design. *Biological Conservation*, **82**, 369–377.
- Kunin, W.E. (1998) Extrapolating species abundance across spatial scales. *Science*, **281**, 1513–1515.
- Kunin, W.E. (2000) Scaling down: on the challenge of estimating abundance from occurrence patterns. *The American Naturalist*, **156**, 561–566.
- Leitner, W.A. & Rosenzweig, M.L. (1997) Nested species-area curves and stochastic sampling: a new theory. *Oikos*, **79**, 503–512.
- Lennon, J.J., Kunin, W.E. & Hartley, S. (2002) Fractal species distributions do not produce power-law species-area relationships. *Oikos*, **97**, 378–386.
- Lennon, J.J., Kunin, W.E., Hartley, S. & Gaston, K.J. (2007) Species distribution patterns, diversity scaling and testing for fractals in Southern African birds. *Scaling Biodiversity* (eds D. Storch, P. Marquet & J. Brown), pp. 51–76. Cambridge University Press, Cambridge.
- May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and Evolution of Communities* (eds M.L. Cody & J.M. Diamond), pp. 81–120. Belknap Harvard University, Cambridge.
- McGeoch, M. & Gaston, K.J. (2002) Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Reviews*, **77**, 311–331.
- McGuinness, K.A. (1984) Species-area relationships of communities on intertidal boulders: testing the null hypothesis. *Journal of Biogeography*, **11**, 439–456.
- Olszewski, T.D. (2004) A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos*, **104**, 377–387.
- Ostling, A., Harte, J. & Green, J. (2000) Self-similarity and clustering in the spatial distribution of species. *Science*, **290**, Technical comment 671a.
- Picard, N., Karambè, M. & Birnbaum, P. (2004) Species-area curve and spatial pattern. *Ecoscience*, **11**, 45–54.
- Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., LaFrankie, J. & Ashton, P.S. (2000a) Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology*, **207**, 81–99.
- Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R. *et al.* (2000b) Predicting species diversity in tropical forests. *Proc. Natl. Acad. Sci. USA*, **97**, 10850–10854.
- Preston, C.D., Pearman, D.A. & Dines, T.D., eds (2002) *New Atlas of the British & Irish Flora*. Oxford University Press, Oxford.
- Rosenzweig, M.L. (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
- Shmida, A. & Wilson, M.V. (1985) Biological determinants of species diversity. *Journal of Biogeography*, **12**, 1–21.
- Sizling, A.L. & Storch, D. (2004) Power-law species-relationships and self-similar species distributions within finite areas. *Ecology letters*, **7**, 60–68.
- Solow, A.R. & Smith, W. (1991) Detecting cluster in a heterogeneous community sampled by quadrats. *Biometrics*, **47**, 311–217.
- Storch, D., Marquet, P. & Brown, J.H., eds (2007) *Scaling Biodiversity*. Cambridge University Press, Cambridge.
- Storch, D., Sizling, A.L., Reif, J., Polechova, J., Sizlingova, E. & Gaston, K.J. (2008) The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. *Ecology letters*, **11**, 771–784.
- Tilman, D. & Kareiva, P.M. (1997) *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton.
- Tjørve, E. (2003) Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, **30**, 827–835.
- Tsao, M. (2000) On the random placement hypothesis. *Environmental and Ecological Statistics*, **7**, 415–422.
- Turner, W.R. & Tjørve, E. (2005) Scale-dependence in species–area relationships. *Ecography*, **28**, 721–730.
- Ugland, K.I., Gray, J.S. & Ellingsen, K.E. (2003) The species-accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888–897.

- Ulrich, W. (2005) Predicting species numbers using species–area and endemics–area relations. *Biodiversity and Conservation*, **14**, 3351–3362.
- Ulrich, W. & Buszko, J. (2003) Self-similarity and the species–area relation of Polish butterflies. *Basic and Applied Ecology*, **4**, 263–270.
- Ulrich, W. & Buszko, J. (2004) Habitat reduction and patterns of species loss. *Basic and Applied Ecology*, **5**, 231–240.
- Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B. & Kunin, W.E. (2004) Spatial patterns in species distributions reveal biodiversity change. *Nature*, **432**, 393–396.

Received 14 January 2008; accepted 22 July 2008  
Handling Editor: Jason Fridley

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Distribution of number of occupied cells for each species in all empirical data sets.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.