

# Geometry of the species–area relationship in central European birds: testing the mechanism

DAVID STORCH\*†, ARNOŠT L. ŠIZLING‡ and KEVIN J. GASTON\*

\*Biodiversity & Macroecology Group, Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK; †Center for Theoretical Study, Charles University, Jilská 1, 110 00-CZ Praha 1, Czech Republic;

‡Department of Philosophy and History of Science, Faculty of Sciences, Charles University, Viničná 7, 128 44-CZ Praha 2, Czech Republic

## Summary

1. The species–area relationship (SAR) is one of the major patterns in community ecology, but the mechanisms that contribute to its exact shape have remained obscure. In continuous mainland areas, the SAR has been attributed to sampling effects (large areas contain species that are too rare to be present in small areas), habitat heterogeneity (large areas contain more types of habitat allowing more species to coexist), and population and metapopulation processes causing spatial aggregation. We tested the contribution of these effects to SARs using data on breeding bird distributions in the Czech Republic, their total population sizes and spatial distributions of their preferred habitats.

2. The relationship between number of species and sampled area is more or less linear on a log–log scale within the Czech Republic, although it reveals saturation when the area is expanded to the whole of central Europe.

3. Neither sampling effect nor habitat heterogeneity alone explain the observed SAR shape: both models predict much higher species richness within any area and a SAR of much lower slope than observed.

4. A combined model based on random sampling constrained by the amount of suitable habitat within an area gives quite realistic predictions of species numbers within different sample areas. Nevertheless, the observed pattern reveals much higher variance of species richness amongst areas, species often being significantly more spatially aggregated than predicted by habitat distribution.

5. Moreover, the relationship between the amount of suitable habitat and the probability of quadrat occupancy is actually nonsignificant for about two-thirds of species, indicating that assumptions of the combined model are unrealistic. Therefore, the shape and slope of SARs are actually affected both by habitat heterogeneity that represents the major driver of distribution of some species, and by spatial aggregation that is not attributable to habitat heterogeneity in other species.

*Key-words:* habitat heterogeneity, spatial distribution, aggregation, metapopulation, power law.

*Journal of Animal Ecology* (2003) **72**, 509–519

## Introduction

The relationship between the number of species and the area sampled is one of the best documented patterns in community ecology (Williamson 1988; Durrett & Levin 1996). Although the fact that larger areas contain

more species than smaller ones is quite obvious, there is no consensus about the exact form of the species–area relationship (hereafter SAR), and the shape and slope of the SAR have remained largely unexplained. The relationship is mostly linear under a log–log transformation, following a power equation (Arrhenius 1921; Rosenzweig 1995), but it may follow an exponential (Gleason 1922; Lennon *et al.* 2001) or logistic equation across some spatial scales (He & Legendre 1996). Moreover, its slope can vary considerably (Connor &

Correspondence: David Storch, Center for Theoretical Study, Charles University, Jilská 1, 110 00-CZ Praha 1, Czech Republic (e-mail: storch@cts.cuni.cz, fax +0420 222220664).

McCoy 1979) and this variability can be partially accounted for by the geographical situation: the slope is higher for isolated areas than for areas nested within one continuous mainland, the highest slope of the relationship being attained by comparison among different continents or other biotic provinces (Rosenzweig 1995). Whereas the differences in species richness among isolated pieces of land are probably strongly affected by the dynamics of colonization/extinction or speciation/extinction (MacArthur & Wilson 1967; Rosenzweig 1995), the SARs within continuous mainland are affected by the factors that determine the spatial distribution of individuals (He & Legendre 1996, 2002).

There are principally three factors related to the spatial distribution of individuals that affect the shape and slope of SARs. The first is the sampling effect: because the majority of species are rare (Preston 1948; Gaston 1994), most will not occur in all of the sampled areas and will be sampled only within larger ones, even if their spatial distribution is random. Therefore the sampling effect itself is capable of producing monotonically increasing SARs (Preston 1962), although it is not sufficient for generating either power-law or realistic slopes of SARs (Leitner & Rosenzweig 1997). The second factor is habitat heterogeneity (Rosenzweig 1995): larger areas host more habitat types, and thus enable coexistence of more species associated with particular habitats. Habitat heterogeneity potentially affects the spatial clustering of individuals, but this can be affected as well by spatial population dynamics, including the dynamics of local colonization and extinction (Hanski & Gyllenberg 1997) or aggregative behaviour (Taylor, Woiwod & Perry 1978). Thus, spatial population dynamics of species may be considered as a third major factor affecting SARs.

The processes contributing to the shape and slope of mainland SARs have been tested only indirectly. Whereas the pure sampling effect could be rejected quite easily, because if the abundance distribution of species is known it gives exact predictions concerning the shape and slope of SARs, testing the other effects (habitat heterogeneity and spatial population dynamics) is complicated because it is not clear which patterns they should produce. For example, an increased number of habitats with area must surely influence SARs, but habitat heterogeneity itself gives no quantitative prediction of their form; it is not clear why habitat heterogeneity should increase with area in such a way that power-law SARs with particular slopes emerge. The effect of heterogeneity has therefore been tested mostly by partialling out area and the diversity of habitats (see Boecklen 1986), with the almost invariant conclusion that habitat diversity indeed correlates with number of species even if area is controlled (Rosenzweig 1995; Gaston & Blackburn 2000). In contrast, there is theory that predicts the quantitative parameters of SARs on the basis of metapopulation dynamics (Hanski & Gyllenberg 1997), but the assumptions do not seem appropriate for many situations

other than the archipelagoes of isolated islands or sufficiently separated and discrete habitat patches (Gaston & Blackburn 2000).

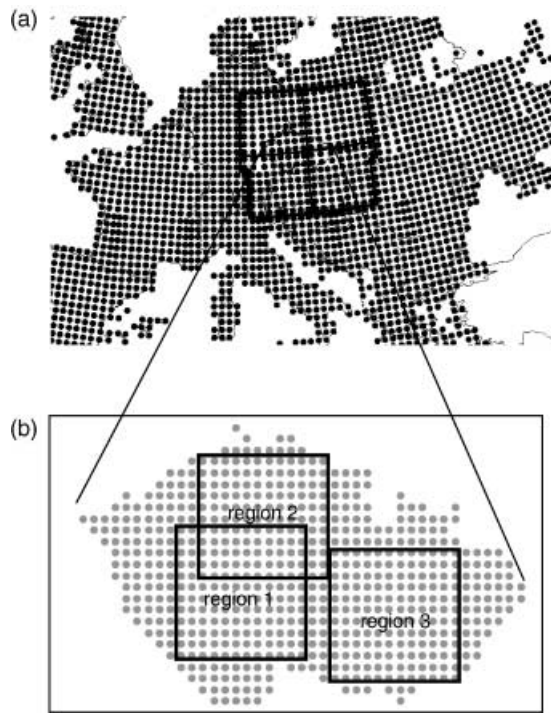
As all of the factors mentioned probably affect the spatial distribution of species, it is not easy to disentangle them and evaluate their importance separately. One way to do this would be to assess the exact mathematical properties of empirical SARs and to compare them with theoretical predictions based on the hypothesized mechanisms (He & Legendre 1996). However, because it is not very clear which theoretical prediction could be derived from particular mechanisms, and because distinguishing individual mathematical forms of the relationship is extremely difficult (Connor & McCoy 1979), this approach is very limited. The second approach consists in building models that include only particular factors of concern, and comparing them with observed SARs (regardless of their exact mathematical form) to assess which of these factors are actually sufficient for producing observed SARs. This approach, however, depends strongly on the availability of data related to the particular mechanisms, i.e. besides data on the real spatial distributions of species, data are also required on population abundances and the spatial distribution of suitable habitats of individual species. As these data are available for birds in the Czech Republic (see Storch & Šizling 2002), we could test to what extent the sampling effect, habitat heterogeneity and population aggregation that is not attributable to habitat heterogeneity are responsible for SARs in this particular situation.

## Materials and methods

### DATA

Data on the breeding distributions of bird species within the Czech Republic were obtained from Št'astný, Bejček & Hudec (1996). Distributions were mapped, in 1985–89, on 628  $12 \times 11.1$  km quadrats; only records of probable or confirmed breeding were included in the analyses. There is a problem with edges of study plots for establishing SARs (Lennon *et al.* 2001) because larger sample areas have to be either located only in the centre of a study plot or must be incomplete. To avoid these difficulties, SARs were established for three large complete square samples (hereafter called regions), each located in a different part of the Czech Republic, and each consisting of 144 ( $12 \times 12$ ) mapping quadrats (Fig. 1b). To assess the form of SARs beyond the extent of the Czech Republic we also used data on the distribution of species in a large square representing the whole of central Europe (Fig. 1a) from Hagemeyer & Blair (1997), which consists of  $16 \times 16$  mapping quadrats ( $50 \times 50$  km each) following the delimitation of that square by Storch & Šizling (2002).

The area of suitable habitat for each species within each mapping quadrat was assessed using the protocol of Storch & Šizling (2002). Thirty-seven land cover



**Fig. 1.** Location of the study plots: (a) the central European square consisting of 256 ( $16 \times 16$ ) mapping quadrats, each of them having area  $2500 \text{ km}^2$  ( $50 \times 50 \text{ km}$ ) and (b) the location of the three regions consisting of 144 ( $12 \times 12$ )  $12 \times 11.1 \text{ km}$  quadrats within the Czech Republic.

types recognized in the CORINE Land Cover Database (based on satellite imagery data) were amalgamated by major structural properties in such a way that the resulting 17 classes represented habitats distinctly occupied by birds. The following habitat classes resulted: Coniferous forests, Deciduous forests, Mixed forests, Water bodies, Large water bodies, Meadows, Swamps and bogs, Heathlands, Orchards and vineyards, Fields, Suburban habitats and villages, Urban habitats, Building sites and other bare ground, Rocks and debris, Shrub and forest regrowth, Open vegetation mosaics, and Large rivers. Each habitat type was assessed as to the suitability for breeding of individual bird species (see Appendix in Storch & Šizling 2002), and for each species the total area of suitable habitats within each quadrat was calculated. Each quadrat was also characterized by minimum and maximum elevation, and only those whose elevational extent overlapped with the breeding elevational extent of species were regarded as suitable. Maximum and minimum estimates of population abundances of species living in the Czech Republic were obtained from Hudec *et al.* (1995). These estimates are conservative (differing often by orders of magnitude for a particular species) which guarantees that the real species abundances almost certainly lie between the extremes.

Any patterns of species diversity can potentially be affected by sampling effort: if this effort increases with sampled area, the probability of detecting a species

also increases, artificially increasing the slope of the observed SAR (Cam *et al.* 2002). This ‘sampling effort effect’ (which we distinguish from the ‘sampling effect’ described earlier) was probably very weak in our data set because all the quadrats were intensively censused for 5 years with the primary aim of proving breeding of all species that actually occurred in a quadrat, and because particular effort was devoted to species with generally lower probabilities of detection (Št’astný *et al.* 1996). However, although this procedure sufficiently accounted for the inherent differences in detection probabilities among species, it is still quite likely that there remained some differences among mapping quadrats, because individual quadrats were censused by different field workers, and thus some could be censused with lower effort. For this reason we used only plots comprising adjacent mapping quadrats that were equal to or larger than the  $2 \times 2$  basic mapping quadrats in those analyses where failure of detection of species within the individual basic quadrats would affect observed patterns. These analyses comprised estimates of slopes of SAR curves, and calculations of species spatial aggregation (see below). The only patterns that remained potentially affected by the sampling effort effect were thus those concerned with the basic mapping quadrats, i.e. the exact shape of whole observed SARs, whose lower part therefore must be interpreted cautiously.

#### MODELS TESTED

We used three models to test which factors are sufficient to explain the observed SARs. These models differed in the constraints imposed on the spatial distribution of individuals, expressed in the probability  $p_{q(A,i)}$  that a breeding pair of species  $i$  will occupy a particular quadrat  $A$ :

*The habitat model* assumed that species occupied all the quadrats that contained suitable habitat ( $p_{q(A,i)} = 1$  for quadrats with suitable habitat, and  $p_{q(A,i)} = 0$  for remaining quadrats). Thus, only habitat composition of quadrats has been accounted for.

*The sampling model* assumed that all species could distribute randomly within all the quadrats ( $p_{q(A,i)} = 1/N$ , where  $N$  is total number of quadrats). The spatial distribution was constrained only by total population abundances, which did not allow the less abundant species to occupy all quadrats.

*The habitat area model* assumed that species colonize quadrats according to the area of suitable habitat within a quadrat. The probability of species’ quadrat occupancy was therefore proportional to the total area of suitable habitat within a quadrat ( $p_{q(A,i)} = P_i/\Sigma P_i$ , where  $P_i$  is the area of suitable habitats within a quadrat and  $\Sigma P_i$  is total area of suitable habitats within the Czech Republic).

## MODEL BUILDING AND TESTING

To allow statistical comparison between predicted and observed SARs we calculated the probabilities of occurrence of particular numbers of species within each sample area. For observed SARs, we simply calculated the frequency of species numbers in all possible plots, i.e. all possible squares containing adjacent basic mapping quadrats within the  $12 \times 12$  square (beginning with the area of one mapping quadrat), and then established median (0.5 quantile) and 95% confidence intervals for each sample area. In the case of the habitat model species were simply replaced by their suitable habitats, and the same procedure was performed. For SARs predicted by the sampling model and the habitat area model we calculated probabilities that a sample area will be occupied by a particular number of species by the following steps:

1. For each plot, we calculated probabilities of presence of each species using the binomial distribution that assumes that species pairs are independent (Coleman 1981; Williams 1995; He & Legendre 2002). It was calculated according to the formula

$$p_{\text{spec}(A,i)} = 1 - (1 - p_{q(A,i)})^{n_i} \quad \text{eqn 1}$$

where  $p_{\text{spec}(A,i)}$  is the probability that the plot  $A$  will contain at least one breeding pair of species  $i$ ;  $p_{q(A,i)}$  is the probability that a breeding pair of the species  $i$  will occupy  $A$  (which differed according to the model used, see above), and  $n_i$  is the number of breeding pairs of the  $i$ -th species.

2. Using these probabilities of occupancy of all possible plots by individual species we calculated the probabilities of occupancy of each plot by a particular number of species  $S$  using the formula

$$P_S = \sum \left[ \prod_{i=1}^s p_{\text{spec}(A,i)} \prod_{j=s+1}^N (1 - p_{\text{spec}(A,i)}) \right] \quad \text{eqn 2}$$

where  $P_S$  is probability of occupancy of a plot by  $S$  species, and  $\Sigma$  is related to all possible combinations of  $S$  species.

This formula is based on the assumption that probabilities of plot occupancy by individual species are independent. Then the probability of occupancy by two species  $A$  and  $B$  from the set of four species  $A, B, C, D$  can be calculated as the probability of occupancy by species  $A$  times the probability of occupancy by species  $B$  times the probability of non-occupancy by species  $C$  times the probability of non-occupancy by species  $D$ . Because of the numerical difficulties and time demands in using this formula directly for all 198 species, a special algorithm was developed (see Appendix). As a result, we obtained the matrix of probabilities of particular  $S$  for each possible plot.

3. The probabilities of particular  $S$  for particular sample areas were obtained as arithmetic means of the probabilities for the set of all plots with a particular area;

this calculation is based on the assumption that each particular plot can be chosen with the same probability.

4. From the resulting matrix of probabilities of each species number  $S$  for each sample area we calculated median and confidence intervals as in the case of observed SARs.

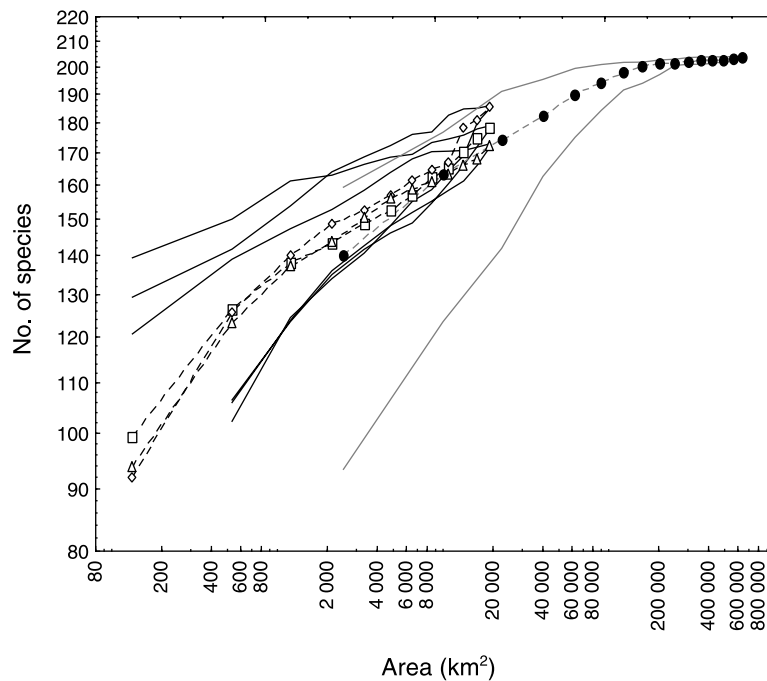
We performed all of these calculations for each region (square  $12 \times 12$ ) in the Czech Republic separately, using firstly the maximum estimates of all species population sizes and secondly the minimum estimates for all species.

The results of the models were compared with observed SARs using the overlap between the confidence intervals of predicted and observed species numbers for each sample area. The significance of the difference between observed and predicted species numbers for each sample area was calculated as a probability that the observed species number of a randomly selected plot will be systematically higher or lower than a randomly chosen species number predicted for that sample area by a given model. The difference between predicted and observed species numbers was then considered as significant when more than 97.5% of randomly selected species numbers from the observed values was either systematically higher or systematically lower than the randomly selected numbers from the predicted distributions.

We also calculated  $z$ , the slope of SAR curves in log-log space (Rosenzweig 1995) for all observations and models (not including plots smaller than  $2 \times 2$  mapping quadrats, see above). Due to the overlapping plots the species richness values for each sample area were not independent, and we could not use standard linear regression to provide reliable confidence intervals. Instead, we used a Monte Carlo method that randomly selected one value of  $S$  for each area, respecting the probability distribution of  $S$  in the case of predicted species numbers. Then we performed linear regression to obtain  $z$  from these random  $S$ -values, and repeated the procedure a thousand times for all models and observations. Median and 95% confidence intervals were then calculated from these randomly generated  $z$ -values.

## EVALUATING RELIABILITY OF ASSUMPTIONS OF HABITAT AREA MODEL

If the habitat area model is correct, birds should preferentially occupy the quadrats with large areas of suitable habitats. Then, the distribution functions of the preferred habitat area within those quadrats observed to be occupied should not differ from these distribution functions for quadrats whose occupancy is predicted by the model. To compare these distribution functions we used a randomization procedure that simulated quadrat occupancy based on the habitat area model. A Kolmogorov-Smirnov test was calculated for the differences in distribution function of preferred habitat area between real occupancy data



**Fig. 2.** Species–area relationships for all three regions within the Czech Republic and the whole of central Europe. Medians for each region are marked as follows:  $\triangle$  = region 1;  $\diamond$  = region 2;  $\square$  = region 3;  $\bullet$  = central Europe. The solid lines represent 95% confidence intervals for species numbers within sample areas.

and 100 simulations (R–S comparison), as well as between each of these simulations and the remaining 99 simulations (S–S comparison). The reliability of the habitat area model was rejected when the average value of Kolmogorov–Smirnov statistics for the R–S comparison was higher than the values for all the 100 S–S comparisons.

We also estimated the level of spatial aggregation of species that is not attributable to the spatial distribution of habitats. We compared the observed number of occupied  $2 \times 2$  quadrats (consisting of 4 adjacent mapping quadrats) with the results of simulations based on the habitat area model with minimum abundance estimates. If the observed number of occupied  $2 \times 2$  quadrats was lower than any number that resulted from all 100 simulation runs, the species distribution was regarded as significantly aggregated.

Finally, we assessed the observed and predicted SARs for the two groups of species: those whose probability of occupancy was significantly affected by the area of suitable habitat, and those for which the habitat area model had been rejected. The significance of differences between predicted and observed species numbers was assessed for each case (see above), and these differences were compared between the two groups of species for each sample area and each region.

## Results

### OBSERVED SARs

The SARs revealed by medians of species number for each sample area within the three regions in the Czech

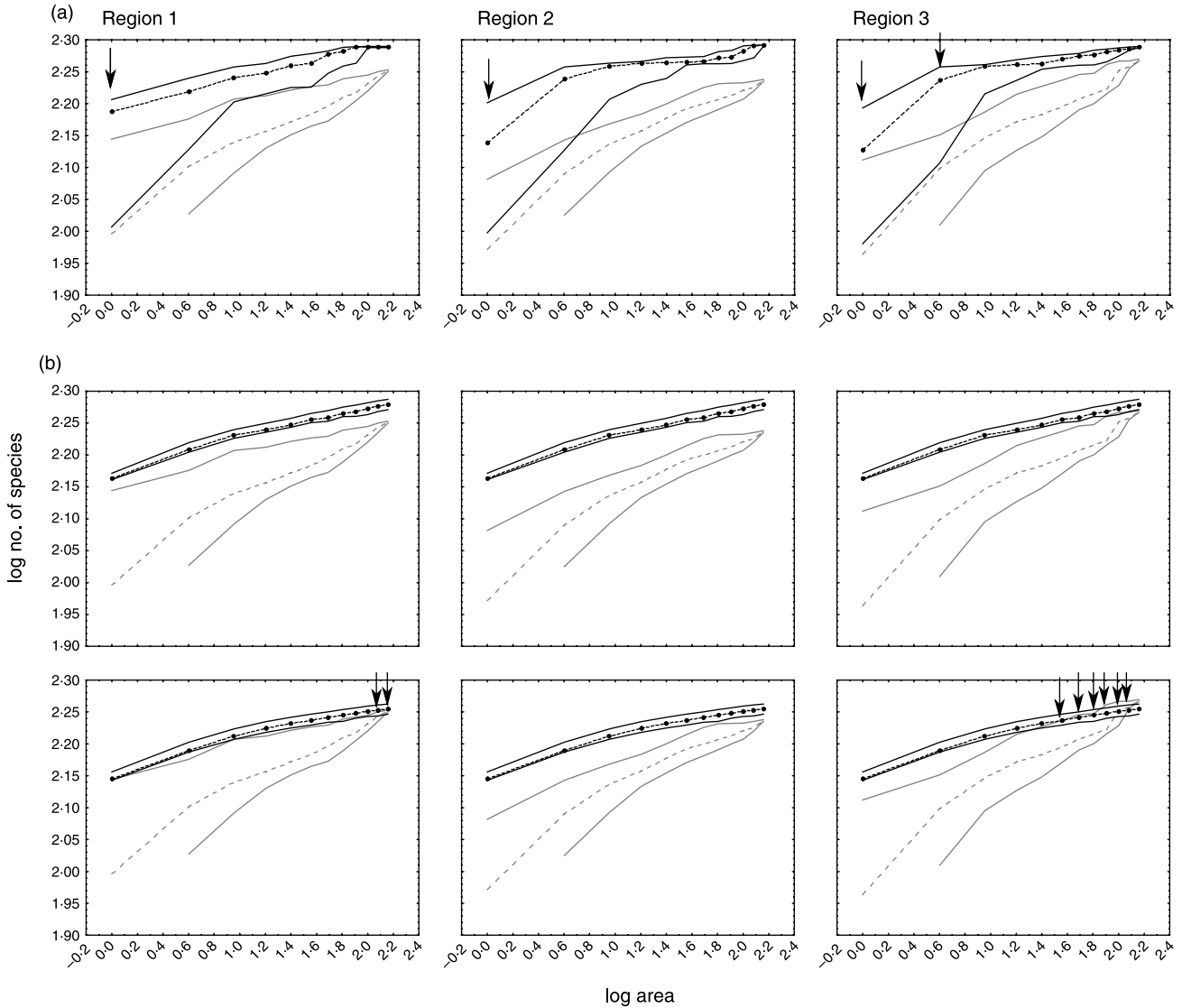
Republic and the single central European square were substantially linear on a log–log scale, but only over particular spatial scales, from *c.* 1000–80,000 km<sup>2</sup>. The slopes of these median SARs across these spatial scales were very similar among all regions:  $z = 0.09$  for region 1,  $z = 0.09$  for region 2,  $z = 0.10$  for region 3, and  $z = 0.09$  for the central European square. The slope was, however, much flatter at the larger spatial scales, strongly indicating saturation within the whole of central Europe (Fig. 2), and seemed to be steeper at the smallest scales (1–4 basic mapping quadrats) due to the relatively low species richness within the basic mapping quadrats. The variance of these slopes, revealed by the confidence intervals (see Table 1), was quite high because different plots of equal area differed substantially in species richness, especially in the case of smaller plots.

### HABITAT MODEL AND SAMPLING MODEL

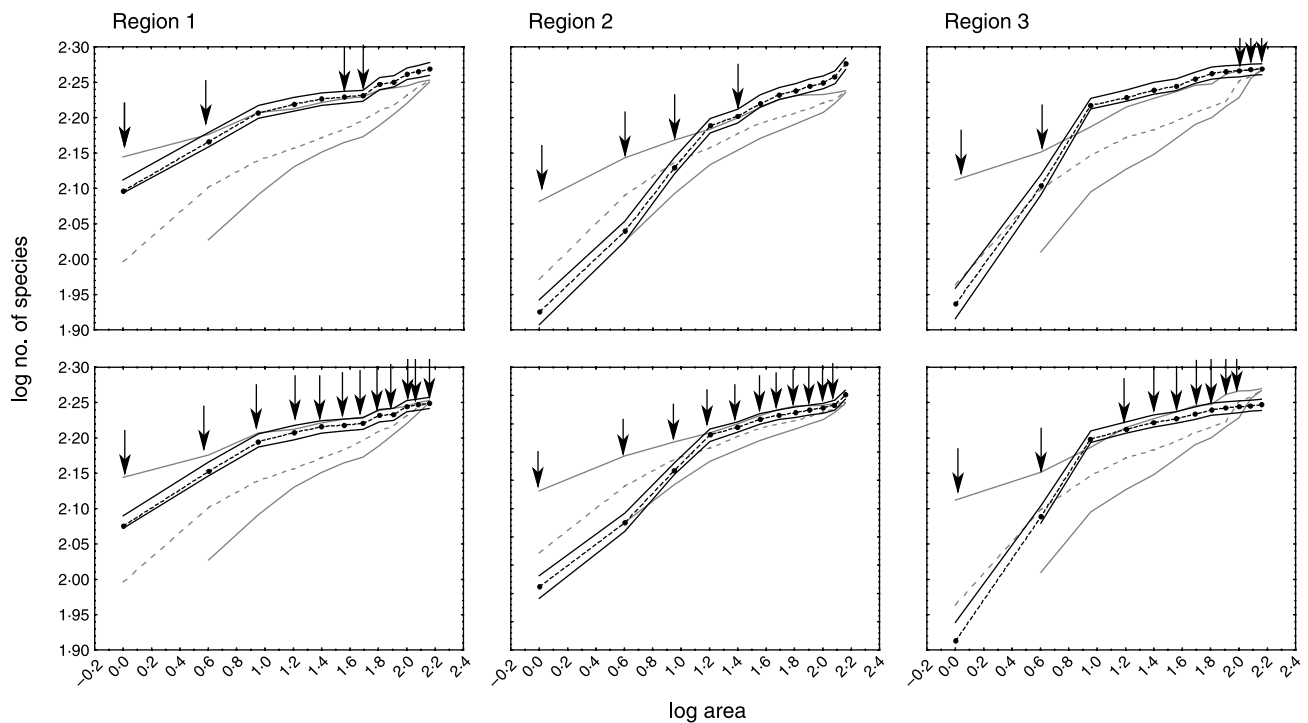
SARs predicted using the number of suitable habitats within samples (habitat model) were very different from those observed (Fig. 3a). The number of suitable habitats increased rapidly with area and quickly became almost saturated, at least for regions 2 and 3, where most suitable habitats were already present in *c.*  $33 \times 33$  km quadrats. The model also predicted much higher numbers of species than observed for all the sample areas. A similar pattern was also shown by the sampling model (Fig. 3b), although the number of species predicted by the sampling effect did not reveal the rapid increase at the beginning, and instead increased very slowly and continually with area (Table 1).

**Table 1.** Statistics (medians and 95% confidence intervals) of slopes of observed and predicted SARs, obtained by Monte Carlo randomization (see Methods). Note that because the habitat area model predicted generally very nonlinear SARs in log–log space, linear regression coefficients do not provide very reliable measures of their shape

	Region 1			Region 2			Region 3		
	Median	–95%	+95%	Median	–95%	+95%	Median	–95%	+95%
Observed	<b>0.090</b>	0.062	0.118	<b>0.087</b>	0.069	0.112	<b>0.103</b>	0.080	0.140
Habitat model	<b>0.046</b>	0.028	0.082	<b>0.033</b>	0.014	0.076	<b>0.035</b>	0.017	0.078
Sampling model (max)	<b>0.043</b>	0.038	0.047	<b>0.043</b>	0.039	0.047	<b>0.043</b>	0.039	0.047
Sampling model (min)	<b>0.039</b>	0.034	0.043	<b>0.039</b>	0.034	0.042	<b>0.039</b>	0.034	0.042
Habitat area model (max)	<b>0.059</b>	0.053	0.064	<b>0.132</b>	0.125	0.138	<b>0.084</b>	0.078	0.091
Habitat area model (min)	<b>0.055</b>	0.049	0.059	<b>0.122</b>	0.115	0.129	<b>0.079</b>	0.073	0.084



**Fig. 3.** Comparison between predicted (black lines) and observed (grey lines) SARs for (a) habitat model and (b) sampling model. The solid lines represent 95% confidence intervals for species numbers within sample areas. In the case of the sampling model, results based on maximum (top) and minimum (bottom) estimates of total species abundances are shown. Arrows indicate those sample areas where the difference between predicted and observed species numbers were not significant, i.e. where observed species numbers fitted those predicted by respective model. Note, however, that the lack of significant differences can be partially due to the high variance of observed species numbers.



**Fig. 4.** Comparison between predicted (black lines) and observed (grey lines) SARs for habitat area model. Results based on maximum (top) and minimum (bottom) estimates of total species abundances are shown. See Fig. 3 for explanation.

The sampling model also predicted much lower variance in the number of species within plots than was observed.

#### HABITAT AREA MODEL

The model based on sampling constrained by suitable habitats, i.e. habitat area model (Fig. 4), also predicted much less variance in species numbers within a particular area than observed, but the predicted numbers fitted well within the confidence intervals of the observed data, especially in the case of minimum estimates of abundances. On the other hand, the exact shape of the SARs predicted by the habitat area model differed markedly from the observed shape, as well as differing among individual regions. The predicted shape remained much more curvilinear and less variable than that observed. Thus, species richness of sample areas is reasonably predictable using the minimum estimates of total population sizes of individual species and knowledge of relative areas of species' preferred habitats within squares, although the exact shape of SARs is not predictable using this information.

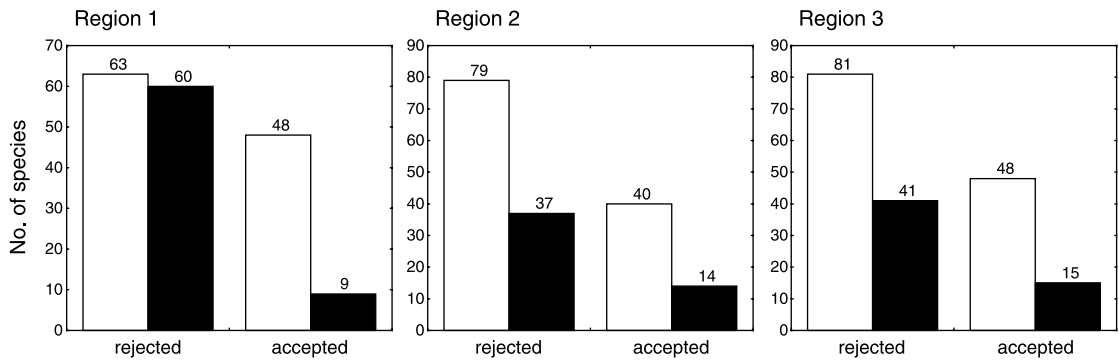
#### RELIABILITY OF THE HABITAT AREA MODEL

The congruence of species numbers predicted by the habitat area model and observed species numbers might not necessarily be caused by the fact that the model represented well those processes really occurring in nature, i.e. colonization of suitable habitat patches according to their areas. The habitat area model would give good predictions of species numbers even if the spatial distribution of species only reached

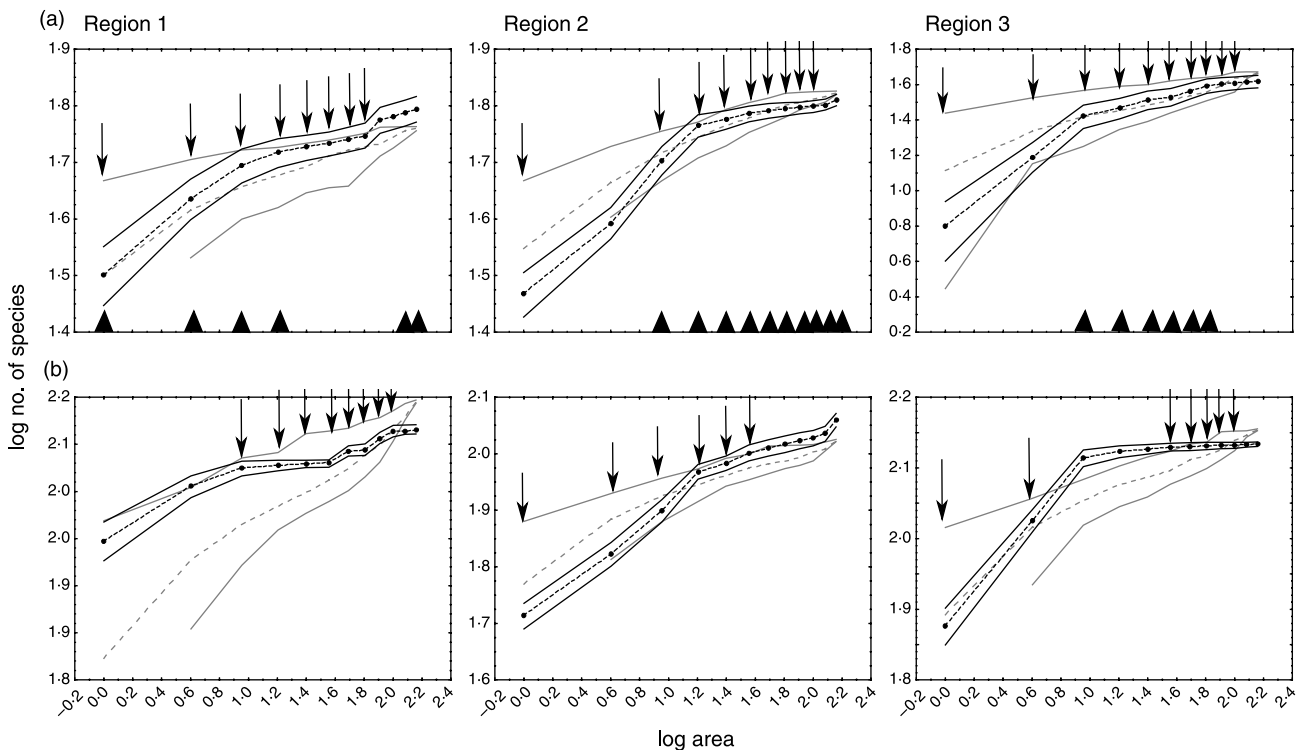
the same level of aggregation as the spatial distribution of habitats, regardless of whether species distribution really matches habitat distribution. In fact, the habitat area model is not reliable for about two-thirds of all the species, and at least one-third of species are significantly more spatially aggregated than predicted by the model (Fig. 5). The habitat area model fitted much more closely to the observed SARs for species whose occupancy was significantly related to the preferred habitat area than for species for which this was not so (Fig. 6). Therefore, the SARs for all species (see Fig. 2) are affected both by habitat heterogeneity that represents the major driver of distribution of some species, and by spatial aggregation that is not attributable to habitat heterogeneity in other species.

#### Discussion

Species–area relationships for birds in the Czech Republic can be expressed as a power law with quite a low slope in comparison to most mainland SARs (cf. Connor & McCoy 1979; Rosenzweig 1995). However, the variability in the number of species within sample areas is very high, and the slope of SARs obtained by regression of median values seems not to be particularly representative. For these reasons it makes no sense to compare different models of SARs, i.e. power vs. exponential functions (Lennon *et al.* 2001) because in any case both functions would represent very rough approximations of the observed relationship. Rather more important was the observation that for the areas larger than those contained within the Czech Republic the SAR became flattened, indicating a saturation in



**Fig. 5.** Numbers of species in which the habitat area model has been rejected and accepted, respectively (see Methods), for the three regions. Black bars represent species whose level of aggregation was significantly ( $P < 0.01$ ) higher than predicted by the habitat; it does not mean that the other species were not aggregated, but that their level of aggregation corresponded to the level of aggregation of the species' preferred habitat (even in the case when their distribution did not follow the distribution of the habitats). This is probably the reason why the number of significantly aggregated species was higher in region 1, which was quite homogeneous in terms of spatial distribution of habitats, and thus significant aggregation was revealed more readily.



**Fig. 6.** Comparison between predicted (black lines) and observed (grey lines) SARs for habitat area model based on minimum estimates of species abundances for two groups of species: (a) those which fitted predictions of habitat area model, and (b) those whose probability of occupancy was not related to the area of suitable habitat. Note that scaling of log number of species differs among the plots because in each case a different group of species fulfilled the criteria. Arrows indicate those sample areas where the difference between predicted and observed species numbers were not significant (see Fig. 3), whereas black triangles indicate the sample areas where the difference between predicted and observed species number in species with rejected habitat area model was higher than in species with accepted habitat area model.

the number of species within large areas. There has been a long discussion as to whether SARs can have an asymptote (He & Legendre 1996; Lomolino 2000; Williamson, Gaston & Lonsdale 2001). Whereas for SARs among isolated islands or provinces there is no theoretical reason or empirical evidence for saturation (Williamson *et al.* 2001), within mainland SARs it has been suggested that it may occur (He & Legendre 1996, 2002). However, our observation is rather excep-

tional (but see Crawley & Harral 2001) and could be attributed to the relative homogeneity of central Europe (Storch & Šizling 2002). This raises the question of the role of habitat in producing SARs.

There is no doubt that a sampling effect alone or a habitat effect alone is not sufficient to explain the relationship between area and number of bird species in the Czech Republic, because both models predicted much higher numbers of species than observed within



all sample areas. This result could potentially be affected by the sampling effort effect if the observed number of species was actually underestimated due to failure in detecting some species, especially in smaller areas (Cam *et al.* 2002). This effect is probably responsible for some portion of the great variability of species number within smaller areas, but not for the whole pattern, because observed numbers were lower than predicted by the models also for large areas, where the sampling effort effect should not play such a role (Cam *et al.* 2002). Moreover, the species numbers predicted by these two models were generally higher than all species numbers documented within any plot of particular area (the confidence intervals did not overlap in the majority of cases), which cannot be attributed purely to sampling effort.

Species numbers thus seem to be constrained by limited availability of individuals (sampling effect) as well as by limited availability of suitable habitats (habitat heterogeneity effect). The combined effect of both factors gives quite reasonable predictions of species numbers within different sample areas. However, it does not seem that these factors are entirely sufficient for explaining the observed shape and slope of SARs. First, the predicted SARs seem to be generally more curvilinear in log–log space than the observed SARs: the power law is not produced by the models and actually emerges in spite of their predictions. Second, although the habitat area model provided reasonable predictions of species numbers, two-thirds of the species do not seem to be distributed according to the assumptions of the model. Spatial distribution of species is not fully attributable to the spatial distribution and size of patches of suitable habitats, because many species do not preferentially occupy quadrats with larger areas of suitable habitats, and the level of aggregation is often even higher than predicted on the basis of habitat.

The higher species aggregation than predicted by habitat could potentially be attributable simply to too broad a habitat delimitation. Species may actually be specialized to some habitats that have not been recognized by the land-cover data, and thus they may be more aggregated if their preferred habitats occur only within some patches of the broadly delimited habitats. Some species are certainly finely specialized (Cody 1985; Storch & Frynta 2000), although there is no reason to expect that their habitats should necessarily be clustered within the broadly delimited habitat types to produce higher spatial aggregation. Moreover, although fine habitat specialization could potentially explain high spatial aggregation, it does not explain the fact that the probability of quadrat occupancy is often not related to the area of the broadly delimited suitable habitat. Species have a tendency to aggregate, but often not in the places with larger areas of suitable habitats, and this observation challenges the role of habitat heterogeneity in producing distributional patterns at least in some species.

But what causes spatial aggregation if not the spatial distribution of suitable habitats? There are several possibilities. Species could settle preferentially in already occupied patches, because the presence of other conspecific individuals serves as a clue to patch suitability (Stamps 1988; Muller *et al.* 1997), because philopatry constrains eventual dispersal, or because suitable habitat patches differ substantially in productivity. Productivity differences among quadrats which are not related to differences in habitat composition are quite probable, as the Czech Republic lies between the Atlantic and continental climatic zones that strongly differ in both precipitation and temperature variability. It is known that productivity-related variables like temperature substantially affect bird distribution (Currie 1991; Lennon, Greenwood & Turner 2000), and thus this factor is likely to affect both species aggregation and the high observed variability of species richness in individual plots.

Current spatial distributions may also not be at equilibrium, some regions being yet uncolonized, and/or some local populations having become extinct. These processes of colonization and extinction may constitute metapopulation dynamics (Hanski 1999), not necessarily implying any equilibrium between colonization and extinction rates. Although there is no evidence that metapopulation structure is particularly common for birds or other taxa (Gaston & Blackburn 2000), it does not mean that processes such as local colonization and extinction do not play a role in determining distributional patterns (Freckleton & Watkinson 2002). Storch & Šizling (2002) showed for the same Czech Republic avian data set that the species whose distributions are unsaturated in terms of number of occupied quadrats are often those revealing decreasing or increasing trends of population change or that occur on the boundary of their geographical ranges. An aggregated distribution is in all these cases almost inevitable.

Even if spatial aggregation is responsible for the observed shape of SARs, it does not answer the question of why the observed SARs are commonly close to a power law, nor why the slope of SARs for mainland areas is mostly between 0.1 and 0.2 (Rosenzweig 1995). We may only hypothesize that because a power law implies scale-invariance (Gisiger 2001), the shape of SARs is a consequence of scale-invariant patterns of spatial aggregation. It is reasonable to expect that a set of species strongly differing in their life-history characteristics will contain species that are aggregated on many spatial scales, and no particular spatial scale will be generally more important than any other scale. Then the power law could represent a null hypothesis of spatial scaling of species richness where all spatial scales contribute equally to SARs. The slope of a SAR could, on the other hand, be affected mainly by scaling of habitat heterogeneity (Williamson 1988; Storch, Gaston & Cepák 2002), as suggested by the congruence of our habitat area model and observed SARs. However,

the relationships between scale-invariance in spatial aggregation and the power-law SAR, as well as between scaling of spatial heterogeneity and the slope of the SAR, remain obscure.

### Acknowledgements

We thank Jana Martinková for her assistance in data management, and Sarah Jackson, Arnold Nagy, Michael Collins and an anonymous referee for useful comments. Ministry of Environment and Agency of Nature Conservation and Landscape Protection of the Czech Republic kindly provided the GIS data of land cover. The study was supported by Grant Agency of Charles University (GUK 106/2000) and by institutional grant Výzkumný záměr CTS (BE MSM 110000001). D.S. was supported by a NATO post-doctoral fellowship, from the Royal Society, London.

### References

- Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.
- Boecklen, W.J. (1986) Effect of habitat heterogeneity on the species–area relationships of forest birds. *Journal of Biogeography*, **13**, 59–68.
- Cam, E., Nichols, J.D., Hines, J.H., Sauer, J.R., Alpizar-Jara, R. & Father, C.H. (2002) Disentangling sampling and ecological explanations underlying species–area relationships. *Ecology*, **83**, 1118–1130.
- Cody, M.L. (1985) *Habitat Selection in Birds*. Academic Press, London.
- Coleman, D.B. (1981) On random placement and species–area relations. *Mathematical Biosciences*, **54**, 191–215.
- Connor, E.F. & McCoy, E.D. (1979) The statistics and biology of the species–area relationship. *American Naturalist*, **113**, 791–833.
- Crawley, M.J. & Harral, J.E. (2001) Scale dependence in plant biodiversity. *Science*, **291**, 864–868.
- Currie, D.J. (1991) Energy and large-scale patterns of animal and plant-species richness. *American Naturalist*, **137**, 27–49.
- Durrett, R. & Levin, S. (1996) Spatial models for species–area curves. *Journal of Theoretical Biology*, **179**, 119–127.
- Freckleton, R.P. & Watkinson, A.R. (2002) Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Journal of Ecology*, **90**, 419–434.
- Gaston, K.J. (1994) *Rarity*. Chapman & Hall, London.
- Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.
- Gisiger, T. (2001) Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biology Reviews*, **76**, 161–209.
- Gleason, H.A. (1922) On the relation between species and area. *Ecology*, **3**, 158–162.
- Hagemeijer, W.J.M. & Blair, M.J. (1997) *The EBCC Atlas of European Breeding Birds*. T. & A. D. Poyser, London.
- Hanski, I. (1999) *Metapopulation Ecology*. Oxford University Press, Oxford.
- Hanski, I. & Gyllenberg, M. (1997) Uniting two general patterns in the distribution of species. *Science*, **275**, 397–400.
- He, F.L. & Legendre, P. (1996) On species–area relations. *American Naturalist*, **148**, 719–737.
- He, F.L. & Legendre, P. (2002) Species diversity patterns derived from species–area models. *Ecology*, **85**, 1185–1198.
- Hudec, K., Chytil, J., Št'astný, K. & Bejček, V. (1995) The birds of the Czech Republic. *Sylvia*, **31**, 97–149.
- Leitner, W.A. & Rosenzweig, M.L. (1997) Nested species–area curves and stochastic sampling: a new theory. *Oikos*, **79**, 503–512.
- Lennon, J.J., Greenwood, J.J.D. & Turner, J.R.G. (2000) Bird diversity and environmental gradients in Britain: a test of species energy hypothesis. *Journal of Animal Ecology*, **96**, 581–598.
- Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966–979.
- Lomolino, M.V. (2000) Ecology's most general, yet protean pattern: the species–area relationship. *Journal of Biogeography*, **27**, 17–26.
- MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton.
- Muller, K.L., Stamps, J.A., Krishnan, V.V. & Willits, N.H. (1997) The effects of conspecific attraction and habitat quality on habitat selection in territorial birds (*Troglodytes aedon*). *American Naturalist*, **150**, 650–661.
- Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- Preston, F.V. (1962) The canonical distribution of commonness and rarity. *Ecology*, **43**, 185–215, 410–432.
- Rosenzweig, M.L. (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
- Stamps, J.A. (1988) Conspecific attraction and aggregation in territorial species. *American Naturalist*, **131**, 329–347.
- Št'astný, K., Bejček, V. & Hudec, K. (1996) *Atlas of Breeding Bird Distribution in the Czech Republic 1985–89*. Nakladatelství a vydavatelství H & H, Praha. (In Czech)
- Storch, D. & Frynta, D. (2000) Evolution of habitat selection: stochastic acquisition of cognitive clues? *Evolutionary Ecology*, **13**, 591–600.
- Storch, D., Gaston, K. & Cepák, J. (2002) Pink landscapes: 1/f spectra of spatial environmental variability and bird community composition. *Proceedings of the Royal Society London B*, **269**, 1791–1796.
- Storch, D. & Šizling, A.L. (2002) Patterns in commonness and rarity in central European birds: reliability of the core-satellite hypothesis. *Ecography*, **25**, 405–416.
- Taylor, L.R., Woiwod, I.P. & Perry, J.N. (1978) The density dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology*, **47**, 383–406.
- Williams, M.R. (1995) An extreme-value function model of the species incidence and species–area relationship. *Ecology*, **76**, 2607–2616.
- Williamson, M.H. (1988) Relationship of species number to area, distance and other variables. *Analytical Biogeography* (eds A.A. Myers. & P.S. Giller), pp. 91–115. Chapman & Hall, London.
- Williamson, M.H., Gaston, K.J. & Lonsdale, W.M. (2001) The species–area relationship does not have an asymptote! *Journal of Biogeography*, **28**, 827–830.

Received 25 October 2002; accepted 24 February 2003

**Appendix**

Calculating the probability of occupancy of a plot by  $S$  species using formula 2.

Let us imagine we have four species with probabilities of occupancy of any quadrat  $p_1, p_2, p_3$  and  $p_4$ . We are asking what is the probability  $P$  of occupancy of a quadrat just by two species.

According to formula 2, the probability can be calculated as

$$P = p_1 p_2 (1 - p_3)(1 - p_4) + p_1 p_3 (1 - p_2)(1 - p_4) + p_1 p_4 (1 - p_2)(1 - p_3) + p_2 p_3 (1 - p_1)(1 - p_4) + p_2 p_4 (1 - p_1)(1 - p_3) + p_3 p_4 (1 - p_1)(1 - p_2)$$

and after multiplying

$$P = p_1 p_2 + p_1 p_3 + p_1 p_4 + p_2 p_3 + p_2 p_4 + p_3 p_4 - 3(p_1 p_2 p_3 + p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4) + 6p_1 p_2 p_3 p_4.$$

After marking the sum of all combinations  $j$  of  $N$  as  $SC_j^N$  and generalizing we can write

$$P_s = \sum_{i=0}^{N-s} (-1)^i C_i^{s+i} SC_{s+i}^N.$$

$SC_j^N$  can be counted sequentially, using following procedure:

$$\begin{aligned} SC_1^4 &= p_1 + p_2 + p_3 + p_4 \\ SC_2^4 &= p_1[p_2 + p_3 + p_4] + p_2[p_3 + p_4] + p_3[p_4] \\ SC_3^4 &= p_1[p_2(p_3 + p_4)] + p_3(p_4) + p_2[p_3(p_4)] \\ SC_4^4 &= p_1[p_2(p_3(p_4))] \end{aligned}$$

The arrows show the additive terms, which can be used for counting the following sum of all combinations  $SC$  (next line) so that three operations need not be performed in the second step ( $SC_2^4$ ) five in the third step ( $SC_3^4$ ), etc.

After generalization, the following procedure has been developed:

$\{SC_s^{NofSpec}$  will be stored into array  $aSC[s]$  for each  $s$  between one and  $NofSpec$ }

**BEGIN**

**for**  $i:=1$  **to**  $NofSpec$  **do**  $aP[i]:=aProbability[i]$ ;

$Np:=NofSpec$ ;

$aSC[1]:=0$ ; **for**  $i:=1$  **to**  $Np$  **do**  $aSC[i]:=aSC[i]+aP[i]$ ;

**for**  $s:=2$  **to**  $NofSpec$  **do**

**begin**

$Np:=NofSpec-s+1$ ;  $aSC[s]:=0$ ;

**for**  $j:=1$  **to**  $Np$  **do**

**begin**

$rSumaP:=0$ ; **for**  $i:=j+1$  **to**  $Np+1$  **do**

$rSumaP:=rSumaP+aP[i]$ ;

$aP[j]:=aProbability[j]*rSumaP$ ;

$aSC[s]:=aSC[s]+aP[j]$

**end**;

**end**;

**END.**