



Species–Area Curves based on Geographic Range and Occupancy

MURIEL NEY-NIFLE*† AND MARC MANGEL†‡

**Laboratoire de Physique, Ecole Normale Supérieure, 69364 Lyon Cedex 7, France*
and †*Department of Environmental Studies and Institute of Marine Sciences,*
University of California, Santa Cruz, CA 95064, U.S.A.

(Received on 8 December 1997, Accepted in revised form on 28 September 1998)

The species–area relationship (SPAR) is one of the cornerstones of ecological science. We use information about the geographic distribution of species to deduce the SPAR from a new model; our results complement existing ones that explain SPARs on the basis of equilibrium theory or species–abundance relationships. We assume that each species is characterized by a geographic range and level of abundance (occupancy). We use a mixture of analytical (in one dimension) and numerical (in two dimensions) methods to create SPARs that are concave and can often be parametrized by a power law with exponent less than 1. We show that the main features of the SPAR depend upon the way that one censuses patches and on the characteristics (range and occupancy) of the species. Our approach identifies the key field variables that need to be measured and have implication for conservation, particularly when one estimates the number of species lost after habitat destruction.

© 1999 Academic Press

1. Introduction

The goal of this paper is to show how species–area (SPAR) and commonality–distance curves are related to biogeographic characteristics of organisms, particularly geographic range and occupancy. To do this, we first develop methods to simulate maps of species living in single regions of different spatial scales. We also briefly consider the case of a collection of disconnected patches, such as a set of islands. Because geographic range and occupancy can be estimated in the field, our theory can be used to show how observed SPAR and commonality can

reveal the spatial arrangement of species and vice versa.

The species–area relationship (SPAR§) is one of the foundations of modern ecological science (Connor & McCoy, 1979; McGuinness, 1984; Rosenzweig, 1995; Durrett & Levin 1996). The commonly used SPAR is that the number of species $S(a)$ contained in area a is

$$S(a) = Ca^z \quad (1)$$

where in the simplest case C and z are constants. Typically, $z < 1$ (observed and also predicted, see e.g. Gotelli & Graves, 1996, Chap. 8), so that although the number of species increases with area, it does so at a decreasing rate [Fig. 1(a)]; this has important implications for conservation, because it means that one predicts only moderate loss of biodiversity associated with considerable

‡Author to whom correspondence should be addressed.
E-mail: msmangel@cats.ucsc.edu

§Following Rosenzweig (1995) we use SPAR rather than SAR.

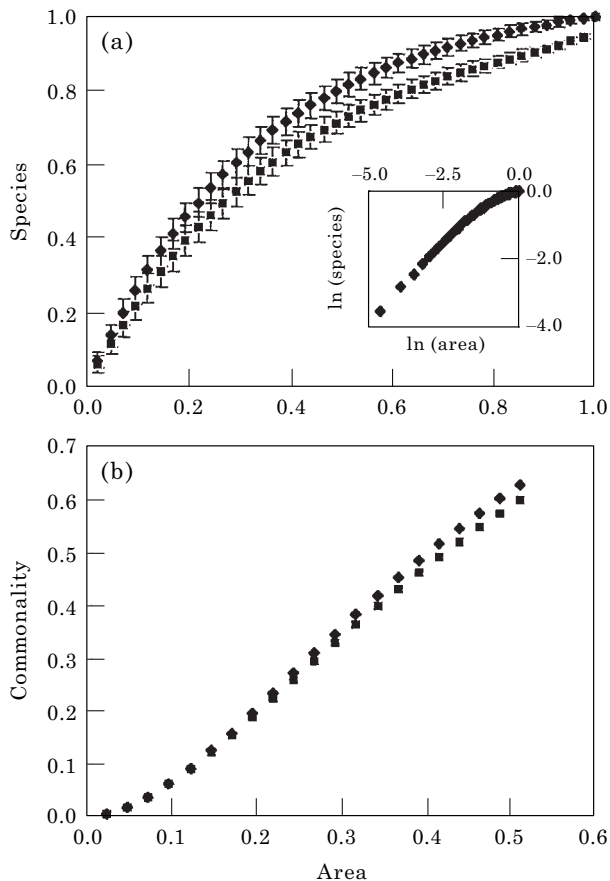


FIG. 1. The SPAR (a) and commonality (b) for species that are distributed according to a uniform occupancy distribution. The maximum occupancy in this community is 0.5% of the total area. Error bars are due to spatial fluctuations. The upper curves are SPARs (a) or commonalities (b) for our model. The lower curves are derived from a continuous version of this model (Coleman *et al.*, 1982). The inset in (a) shows a \ln - \ln plot of our model result. A linear fit to this curve [which corresponds to a power law fit to the original curve, see eqn (1)] gives $z = 0.7 \pm 0.1$. *Exact results.*

loss of habitat (e.g. Simberloff, 1992; Koopowitz, 1994; Holland & Petersen, 1995).

The usual procedure to determine the SPAR at a single field site is to count species in patches of increasing area, and eventually average the number over different locations. Another approach is to determine the abundance curve (number of individuals of each species), and get the SPAR from a model that links the two (Preston, 1962; Coleman, 1981; Coleman *et al.*, 1982). Here we propose to assess the geographic ranges of species (e.g. Legendre & Fortin, 1989

or Mauer, 1994), and deduce the SPAR with a new model.

Common explanations for the SPAR based on equilibrium theory (MacArthur & Wilson, 1967) or species-abundance distributions (Preston, 1962; Coleman, 1981) assume that each species is uniformly distributed in space, these explanations are subject to controversy (see e.g. Gotelli & Graves, 1996). Furthermore, it has been argued that spatial patterns play a role for some communities (Leitner & Rosenzweig, 1997), and one should consider environmental conditions such as habitat diversity when constructing SPARs (Buckley, 1982; McGuinness, 1984; Gotelli & Graves, 1996). Here, we show how the geographic arrangement of species affects the SPAR. We describe the spatial organization of a community with two probability distributions, one for the geographic range (area inside a curve that bounds occupied sites) and one for occupancy (actual area occupied by the species, i.e. the geographic range minus open spaces between occupied sites). Our approach produces SPARs that vary significantly in shape according to the specifics of these distributions. However, they share features that we specify. We also generate commonality curves (species in common to two spatially distinct patches).

In Section 2, we describe the model and two quantities of interest, namely the species-area relationship and the commonality. In the next three sections, we present several calculations in various cases that are then put together in the discussion section. We first consider two types of species that bound the members of a community living on a single area. In Section 3, we consider a widespread species community, in which all members have a minimal degree of aggregation (random placement of species). In Section 4, we consider aggregated species, in which all members have a maximal degree of aggregation, i.e. range is a single patch. The SPARs deduced in this manner will thus bound a more realistic SPAR for a community that mixes species with various degrees of aggregation. In Section 5, we use a simple assumption to determine island SPARs that are compared with nested SPAR in Section 6. Section 7 is a summary of the results, where we consider potential applications of the model.

2. The Model

2.1. GENERATING MAPS OF SPECIES

We are going to show how SPARs can be constructed knowing the spatial arrangements of species. If one draws a map of a region with colors representing various species, one can see patterns, e.g. patchy superimposed colors. Some species can be found everywhere and have a density ranking from rare to common. Other species are aggregated in one or several patches (see e.g. Gaston, 1990). These patterns reflect environmental heterogeneity (in habitat, resources, altitude ...) and interspecific interactions; they can vary through time. In order to take into account spatial heterogeneity without specifying the system in detail, we proceed as follows. A grid of cells is superimposed on this map. Each cell can support one or more species. We count the number of species in each cell and specify their presence or absence.

The simplest assumption one can make about the spatial arrangement of species is that they are uniformly and independently distributed in space. This is called the passive sampling hypothesis (see e.g. McGuinness, 1984). It is a zero order approximation and is the starting point of our analysis. Each species is assigned a number of cells, namely its *occupancy* (Gaston, 1994), which is drawn from a random distribution. Different occupancy distribution functions lead to different SPARs. With the assumption of uniform species distributions, there is no variation through space apart from statistical ones. If a cell is restricted to hold one and only one individual, then occupancy is equivalent to the abundance of the species. Furthermore, in the continuous limit of our model, that is, when the cell size tends to zero, one recovers the random placement model (Coleman, 1981; McGuinness, 1984; Gotelli & Graves, 1996). It has been studied in detail by Coleman who calculated the SPAR averaged over the abundance distribution, and the variance. We will compare our results to his.

As a first order approximation, we assume that some species are aggregated. Beside occupancy, each species will be assigned a *geographic range*. Aggregation is maximum for a highly restricted geographic range, i.e. species occupying only one

small patch. This case was investigated in continuous space by Leitner & Rosenzweig (1997). The other limiting case, i.e. minimum aggregation, corresponds to widespread species, i.e. species whose range is the total area.

Given a number of species and a map divided into cells, a configuration of species is constructed as follows. Each species is characterized by an occupancy (number of cells occupied), n_i , and geographic range (area within a curve that bounds all occupied sites, including empty sites between occupied patches), r_i , both picked randomly from appropriate distributions. Occupancy is smaller or equal to the range. The two can be correlated: some empirical data suggest proportionality (Gaston, 1994), while Leitner & Rosenzweig (1997) use a power law relationship. Species maps are drawn independently and then superimposed. We do not study interactions between species specifically here, but our methods are easily adapted to study guild or antagonist communities. Since the species do not have necessarily comparable spatial requirements, we can deal with a community of species from different taxonomic groups as long as one can assume a single distribution for the geographic range and one for the occupancy.

2.2. CHARACTERIZING SPATIAL ARRANGEMENT OF SPECIES

For a given map of species, several SPARs can be computed depending on the way areas are censused. One could start with one center, say in the middle of the region, count the species in a patch, say a square, and enlarge the square until the border of the region is reached. Within the passive sampling hypothesis (all species are uniformly distributed), going from one center to another, or even changing the shape of the patch considered, will not significantly change the SPAR. This is not true when some species are geographically restricted. Therefore the SPAR will be either averaged over many centers (simulations) or averaged over the distributions $P(r_i)$ (analytical calculation). The analogy is to phase space and time averages in statistical physics, but we know of no ergodic theorem in this case (Rosenzweig, 1998).

The shape of the SPAR can sometimes discriminate between communities, and reveals

the main features of the distributions $P(r_i)$ and $P'(n_i)$. But this is not true in general. There are other quantities that might be used to distinguish between communities having uniformly distributed species and communities having some aggregated species. Consider the number of species in common to two patches of equal area at a distance d . We call it the commonality (Harte *et al.*, 1997a, b, use commonality for this quantity divided by the number of species in the areas considered). The commonality is analogous to the variogram (Cressie, 1991) since it describes spatial correlation and has been measured in some field studies (Whittaker, 1972; Cody, 1993; Palmer & White, 1994; Harte *et al.*, 1997b).

3. A Community of Widespread Species

First, we assume that all members of the community are widespread in the sense that each species is distributed randomly on the grid. In the terminology that we have introduced in the previous section, this means that geographical ranges cover the whole area under consideration. Therefore, the spatial arrangement of the species is entirely specified by the occupancy distribution.

3.1. ANALYTICAL RESULTS

We start with a region of area A populated by a pool of N species (this can be any area with any shape or a collection of patches like islands in a lake). Smaller areas, denoted a , are defined as collections of unit cells, either connected or not. Each species is assigned an occupancy consisting of n_i occupied cells randomly placed. Occupancies are drawn from a single distribution function, $P'(n_i)$. We will show that this model has a continuous limit in which one recovers the random placement model studied by Coleman (1981).

The probability that the first cell designated to be occupied by species i is not in area a is $(A - a)/A$; if all subsequent areas are unit valued, then the probability that it is not found in the second cell is $(A - a - 1)/(A - 1)$, etc. . . . Thus the probability that species i is

not present in a set of cells of total area a is

$$q_i(a) = \prod_{m=0}^{n_i-1} \frac{A - a - m}{A - m} \quad (2)$$

if $n_i \leq A - a$ and $q_i = 0$ otherwise. The probability that this species is present in a is denoted by $p_i(a) \equiv 1 - q_i(a)$. In the limit where the unit cell much smaller than A , eqn (2) simplifies to

$$q_i(a) \approx \left(1 - \frac{a}{A}\right)^{n_i} \quad (3)$$

if $n_i \ll A - a$ [i.e. m is neglected in eqn (2)]. This is exactly what Coleman obtained with the following model. Each species is assigned a number of individuals that are randomly placed in area A . In particular, the probability that a given area a supports all individuals of a given species might be small but not zero. Here, an occupancy of n_i cells translates into a number of individual of n_i .

In general, the average number of species in a is

$$\begin{aligned} \bar{S}(a) &= \sum_{i=1}^N p_i(a) = N - N \sum_{n=1}^{A-a} P(n) \prod_{m=0}^{n-1} \\ &\times \frac{A - a - m}{A - m} \end{aligned} \quad (4)$$

so that

$$\begin{aligned} \frac{d\bar{S}}{da} &= N \sum_{n=1}^{A-a} P(n) \sum_{k=0}^{n-1} \frac{1}{A - k} \prod_{\substack{m=0 \\ m \neq k}}^{n-1} \\ &\times \frac{A - a - m}{A - m} \end{aligned} \quad (5)$$

decreases monotonically as a increases, implying that the SPAR has a concave shape regardless of the distribution $P(n)$. The variance is (see Coleman, 1981; and Feller, 1950)

$$\text{Var}(S) = \overline{(S - \bar{S})^2} = \sum_{i=1}^N p_i q_i \quad (6)$$

Now we count the number of species, $C(a_1, a_2, d)$, in common to two patches of areas a_1

and a_2 at distance d from each other. This quantity does not depend on the specific patches or the distance d , but only on areas a_1 and a_2 . The probability that species i is present in *both* a_1 and a_2 is

$$p_i(a_1, a_2) = 1 - q_i(a_1) - q_i(a_2) + q_i(a_1)q_i(a_2|a_1) \quad (7)$$

where the probability that species i is not in a_2 knowing that it is not in a_1 is

$$q_i(a_2|a_1) = \prod_{m=0}^{n_i-1} \frac{A - a_1 - a_2 - m}{A - a_1 - m} \quad (8)$$

In the limit, where “ i in a_1 ” and “ i in a_2 ” are independent events, as in Coleman’s model, eqn (7) simplifies to $p_i(a_1, a_2) = p_i(a_1)p_i(a_2)$.

In general, the commonality is given by

$$\bar{C}(a_1, a_2, d) = \sum_{i=1}^N p_i(a_1, a_2) \quad (9)$$

which, in the case with $a_1 = a_2 = a \leq A/2$ is

$$\bar{C}(a, a, d) = N \left[1 + \sum_{n=1}^{A-a} P(n) \left(\prod_{m=0}^{n-1} \frac{A - 2a - m}{A - m} - 2 \prod_{m=0}^{n-1} \frac{A - a - m}{A - m} \right) \right] \quad (10)$$

Note that since the commonality does not depend on the distance between the two patches, it is also $2\bar{S}(a) - \bar{S}(2a)$.

Since our model becomes identical to the random placement model of Coleman when $n \ll A - a$, we expect similar curves for $S(a)$ or $C(a, a, d)$ until a approaches $O(A - n_{\max})$.

If $P(n)$ is known, one can derive $S(a)$ and $C(a, a, d)$. Conversely, knowing S and C permits one to determine the moments of the distribution $P(n)$ (Coleman, 1981). To investigate this correspondence, we use a uniform distribution for geographical range. Assuming that species occupancies take discrete values between n_{\min} and n_{\max} ,

$$P(n) = \frac{1}{n_{\max} - n_{\min}} \quad (11)$$

if n is between the two limits, and $P(n) = 0$ otherwise. Using this expression in eqns (4, 6, 10) and $n_{\min} = 1$, we find that if $n_{\max} = A/200$ ($A = 1000$, so n_{\max} is only 5), the SPAR is close to a power law [Fig. 1(a)]. Geographic ranges are small, so species are relatively localized. By contrast, when $n_{\max} = A/10$, the SPAR grows linearly and saturates rapidly at its maximum value [Fig. 2(a)] reflecting the fact that species are distributed widely.

One can compare these curves with those for the random placement by taking the small- n limit. The summations in the various equations are evaluated in the continuous limit. Straight-forward integrations yield the average number of species in a . When $n_{\min} = 0$, we obtain

$$\bar{S}(a) = N \left[1 - \frac{(1-a)^{n_{\max}} - 1}{n_{\max} \ln(1-a)} \right] \quad (12)$$

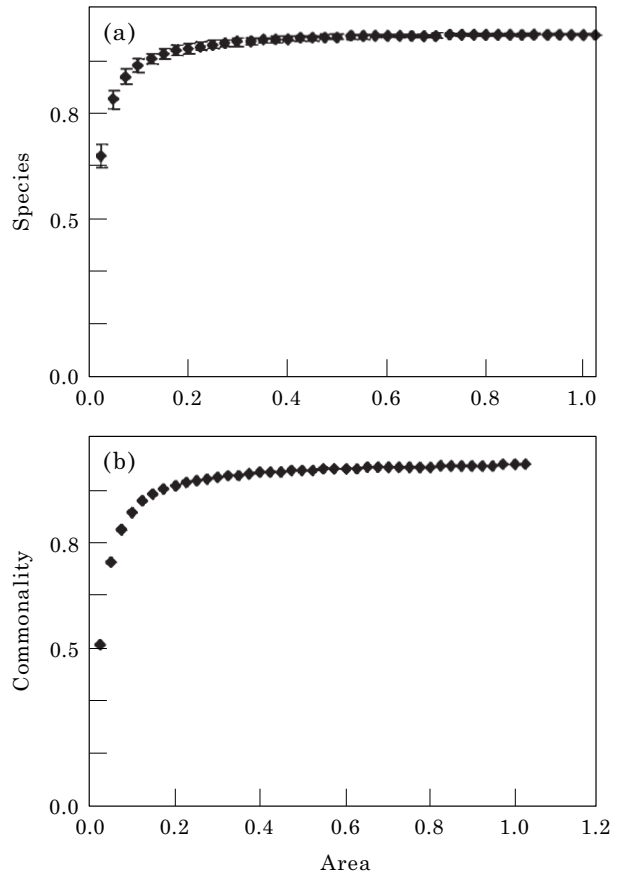


FIG. 2. The SPAR (a) and commonality (b) derived from our model for uniformly distributed species, with a uniform occupancy distribution between 0 and 10% of the total area. *Exact results.*

where

$$a = \frac{a}{A} \quad (13)$$

And the number of species in common to two patches having the same area is

$$\begin{aligned} \bar{C}(a, a, d) \\ = N \left[1 - \frac{4(1-a)^{n_{\max}} - 3 - (1-a)^{2n_{\max}}}{2n_{\max} \ln(1-a)} \right] \end{aligned} \quad (14)$$

Figure 1(a,b) show the similarity of the curves for the discrete model and its continuous version.

For some communities, the geographic range distribution might be well approximated by a gamma density such that most species have small ranges and relatively few have large ones (Gaston, 1990):

$$P(n)dn = \frac{n^{k-1} e^{-n/\omega}}{\Gamma(k)\omega^k} \quad (15)$$

Depending on the two parameters, k and ω , this function has a maximum or not, and is asymmetric (the mean and the variance of this distribution are, respectively, $k\omega$ and $k\omega^2$). In this case, the continuous limit gives

$$\bar{S}(a) = N \left[1 - \frac{1}{(1 - \rho\omega)^k} \right] \quad (16)$$

where

$$\rho \equiv \ln(1-a) \quad (17)$$

This was first derived by Coleman (1981).

To characterize the shape of the SPAR given by eqn (16), we use the slope in the small area limit

$$\frac{d\bar{S}}{da} \approx \frac{N}{A} k\omega \quad (18)$$

and the area at which 90% of the species are included [set $\bar{S} = 0.9N$ and get a] from eqns (16, 17)]

$$a_{90} = 1 - e^{(1 - 10^{1/k})/\omega} \quad (19)$$

These show that unless the fraction of species with small range is large (i.e. the average range, k , is about 0.1% of the total area), the SPAR saturates very quickly. It is more concave than a power law (Fig. 3). We plot the various

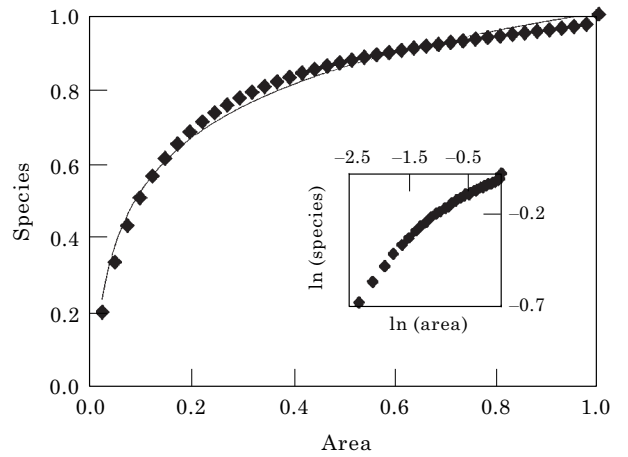


FIG. 3. The SPAR for a community of uniformly distributed species, with exponential occupancy distribution favoring very small values (continuous model). A linear fit to the \ln - \ln plot in inset gives $z = 0.3 \pm 0.1$, but a better fit to the SPAR is $S(a) = 0.2 \ln(a) + 1$ (the continuous line) rather than a power law. *Exact results.*

quantities for $k = 1$ and $\omega = 10$ [in eqn (15)] so that $P(n)$ decreases exponentially as n increases. In that case, eqns (18, 19) imply that the slope at the origin is $10N/A$ and that 90% of the species are included when a/A passed 0.6 (Fig. 3). In reality, in most field studies one does not reach this saturation point, which is one reason why it is difficult to predict the total number of species. These results are similar to those with the uniform distribution of eqn (11). In Section 6 we discuss how these conclusions contrast with results for a community of aggregated species derived in Section 4.

3.2. NUMERICAL SIMULATIONS

Numerical simulation allows one to test a wider range of occupancy distributions. For each distribution, we pick N cells to be occupied, and then distribute them uniformly on the grid. We repeat this procedure a number of times to reduce sampling error (which is smaller than the data point symbols on all our figures). We plot SPAR averaged over the various possible locations of the censused area [from eqn (4)], and the variance gives the error bars [from eqn (6)].

The SPAR can be fitted by a power law, eqn (1) with exponent smaller than one, only if the occupancy distribution is bounded at very low values (typically no species may occupy more than 5% of the total area), like in Fig. 1(a). As



FIG. 4. The Kawasaki algorithm begins with a random distribution of occupied (black) cells, and creates aggregated distributions as Monte Carlo time increases. We show two cases, corresponding to different occupancies: 10% (a) and 50% (b), and three Monte Carlo time steps. At each step, one considers A (total number of cells) pairs of neighboring cells for an exchange (see Appendix). The lower figure required 300 000 steps. *Simulation*.

the number of species with large occupancy grows, the SPAR becomes more concave than a power law, with a rapid growth towards saturation [Fig. 2(a)].

We also consider two communities with increasing degree of aggregation, but identical occupancy distributions: (i) uniformly distributed species, (ii) aggregated species in uniformly distributed patches of various sizes (but isotropic shapes). The “patchy arrangements” are generated with a method derived from statistical physics (Appendix). In this simulation, computing time plays the role of aggregation (see Fig. 4). That way, we study the effect of cell size since one can consider that a patchy arrangement is a uniform distribution on

a grid with larger unit cells. In case (ii), the SPAR and commonality are computed as follows. We randomly pick a set of centers, and squares of increasing size are defined at these centers. The number of species in each square of the same size is determined, and then an average over the various centers is performed. The SPAR for both cases are not significantly different [Fig. 5(a)] indicating that we can classify both patterns as widespread species communities (also suggested by Williams, 1995).

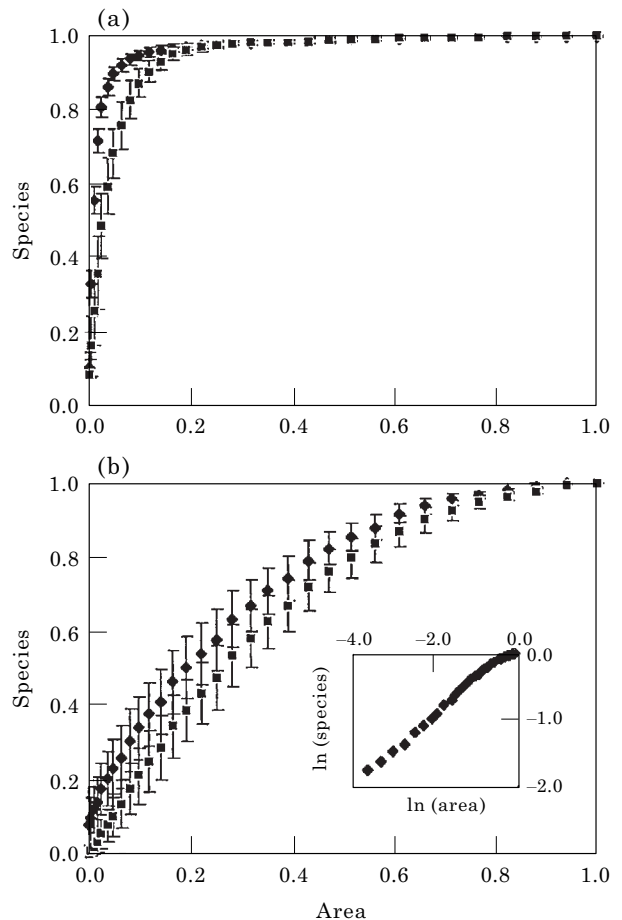


FIG. 5. (a) Widespread species with different degrees of aggregation (upper curve: non-aggregated, and lower curve: patchy) give similar SPARs. The occupancy distribution is uniform between 0 and 20% of the total area. (b) A restricted community also shows similar SPARs with different densities (upper curve: density 1, and lower curve: density 0.01) within a patch. The range distribution is a truncated exponential and vanishes at about half of the total area. The SPAR can be approximated by a power law with exponent $z = 0.5 \pm 0.05$ (see \ln - \ln plot in inset). *Simulation*. Error bars due to sampling of the distributions are smaller than the data point symbols in every simulation (sample size is 50).

4. A Community of Geographically Restricted Species

Now we consider a community in which each species is localized in its own patch. We first consider the case where occupancy equals geographic range, r . The shape and the size of each species patch can vary, and the SPAR might depend on the shapes of censused areas, a . To simplify, we study regular isotropic patch areas, namely squares (or intervals in one dimension), which are distributed uniformly in the region of interest. More realistic patch shapes are studied in Section 4.2. We show that as long as patches are isotropic and randomly distributed, any shape leads to the same SPAR.

4.1. ANALYTICAL RESULTS

We study a particular spatial discretization where analytical calculations are possible. If one considers regions with a gradient, such as altitude, one can expect that sometimes species arrange themselves in strips perpendicular to the gradient. In this case, species can be counted in one direction, which allows for a one-dimensional model.

We assign the N species to an interval of length A that is divided into cells (unit intervals). For each species a geographic range, r_i , is picked from the distribution $P(r_i)$. Then one chooses an interval of length r_i at random and assigns it to species i . Occupancy is $n_i = r_i$ (i.e. density = 1 within a patch). This is done independently for all species. The probability, p_i , that species i belongs to a patch of length a depends on the location of a . We consider two typical cases. First, if a is adjacent to one edge then

$$p_i(a) = \frac{a}{A - r_i} \quad (20)$$

if r is not larger than the remaining interval (i.e. $r_i < A - a$), and $p_i(a) = 1$ otherwise. Second, if a is in the middle of the total interval A , then

$$p_i(a) = \frac{a + r_i}{A - r_i} \quad (21)$$

if r is not larger than the remaining area on the side of a (i.e. $r_i < (A - a)/2$), and $p_i(a) = 1$ otherwise.

We determine the SPAR in the two corresponding cases, which will be of importance for the discussion on *edge effects*. First, areas censused are all adjacent to the same edge and the SPAR is given by

$$S_0(a) = N \sum_{r=1}^{A-a} P(r) \frac{a}{A-r} + N \sum_{r=A-a+1}^A P(r) \quad (22)$$

Second, areas are in the middle of the interval A , so that

$$S_c(a) = N \sum_{r=1}^{(A-a)/2} P(r) \frac{a+r}{A-r} + N \sum_{r=(A-a)/2+1}^A P(r) \quad (23)$$

We talk of an edge effect when eqns (22) and (23) give different results (we will see that, with some range distributions, this means that species are less abundant near the edges).

In order to avoid this edge effect, one could also consider periodic boundary conditions (that is, a loop in dimension one or a torus in dimension two). Then eqn (23) would hold for any interval a along this loop. Leitner & Rosenzweig (1997) do a similar calculation but, in a continuous two-dimensional space. On a two-dimensional grid with periodic boundary conditions and square ranges, the number of species is $S_{2d}(a) = N \sum_r P(r) p_{r,a}$ where $p_{r,a}$ is the probability that range r overlaps area a (denoted p_i above).

One obtains this probability by looking for all possible locations of a species whose range is a square of area r such that it overlaps a given square of area a (just as we did above), in one dimension. A species with range r larger than $(\sqrt{A} - \sqrt{a})^2$ will overlap a given area a , i.e. $p_{r,a} = 1$, and $p_{r,a} = (\sqrt{r} + \sqrt{a})^2$ otherwise [see also eqn (2) in Leitner & Rosenzweig, 1997]. Therefore, the number of species can be written as a sum of two terms, similar to eqn (23), which can be evaluated knowing the range distribution.

Returning to the one-dimensional case, the number of species in common to two patches of length a_1 and a_2 at a distance d is

$$C_c(a_1, a_2, d) \approx N \sum_{r=d}^A P(r) \frac{r-d}{A-r} \quad (24)$$

where $(r - d)/(A - r)$ is the probability that range r contains the interval of length d that separates a_1 and a_2 . To avoid edge effects we consider areas a_1 and a_2 that are far from the edges (which is indicated by the subscript C). The distance d is limited by the size of the total interval and of a_1 and a_2 . As d increases, C decreases and vanishes when $d = r_{\max}$. Furthermore, the intercept when $d = 0$ (adjacent patches) is a constant that depends on the distribution $P(r)$, since $C_c(a, a, 0) = 2S_c(a) - S_c(2a)$.

One could use any distribution and determine the expressions for S and C numerically. Here we consider the uniform distribution for comparison with the results of the previous subsection. Combining eqns (22–24) and (11), we perform the summation. The results depend on how r_{\min} , r_{\max} and $A - a$ are ordered.

When $r_{\min} = 0$ and $r_{\max} = A$, for areas adjacent to one edge (with $a \equiv a/A$)

$$S_0(a) = N(a - a \ln a) \quad (25)$$

which gives

$$\frac{dS_0}{da} = N \ln \frac{A}{a} \quad (26)$$

On the other hand, for areas centered in the middle,

$$S_c(a) = N \left[a - (a + 1) \ln \left(\frac{1 + a}{2} \right) \right] \quad (27)$$

from which

$$\frac{dS_c}{da} = N \ln \frac{2A}{A + a} \quad (28)$$

Thus, when $a/A \ll 1$, $S_0 = 0$ whereas $S_c = N \ln(2)$ [the latter depends on the distribution $P(r)$ in the general case]. The slope of S_c at small areas is steeper than the one of S_0 , but in both cases the slope of the SPAR decreases as a increases and vanishes when a approaches A . Thus, the SPAR has a concave shape and is asymptotic to N , regardless of the distribution $P(r)$. In general $S_c(a) - S_0(a) > 0$ and the difference increases with r_{\max} , i.e. the edge effect.

For $r_{\min} = 0$ and $r_{\max} \ll A$ (so that we consider commonality without the edge effect), the

number of species in common to two patches of area a at a distance d is

$$C_c(a, a, d) = N \left[\frac{d}{r_{\max}} - 1 - \left(\frac{A - d}{r_{\max}} \right) \ln \left(\frac{A - r_{\max}}{A - d} \right) \right] \quad (29)$$

where $d < r_{\max}$. Since

$$\frac{\partial C}{\partial d} = \frac{N}{r_{\max}} \ln \left(\frac{A - r_{\max}}{A - d} \right) \quad (30)$$

$|\partial C_c / \partial d|$ decreases as r_{\max} increases.

4.2. NUMERICAL SIMULATIONS

One is forced to use numerical simulations to study a two-dimensional map and various geographic range distributions. A map of species is constructed as follows: one picks N areas from the range distribution, and for each value assigns a patch of the given area to the species (since patches have a fixed shape, they can not be located anywhere). Patches are either squares or isotropic (see Fig. 4). The SPAR is not significantly different in either case. The critical assumptions are the fact that patches are isotropic and randomly distributed in space (in both cases) which induce an edge effect: there are fewer species near the edges than in the middle of the map. The edge effect is shown analytically in one dimension in Section 4.1, and numerically in two dimensions (Fig. 6). It is an artifact of the simulation, as noted by Colwell & Hurtt (1994). The larger the species ranges, the deeper the edge effect penetrates into the region.

As mentioned above, one way to suppress the edge effect is to use periodic boundary conditions, that is, place species on a torus (Leitner & Rosenzweig, 1997). The expression for the number of species given in Section 4.1 [between eqns (23) and (24)] can be evaluated directly when the range distribution is a gamma density, eqn (15), and summations are converted into integrals. We can also use our numerical simulation; we did both as a test of this simulation.

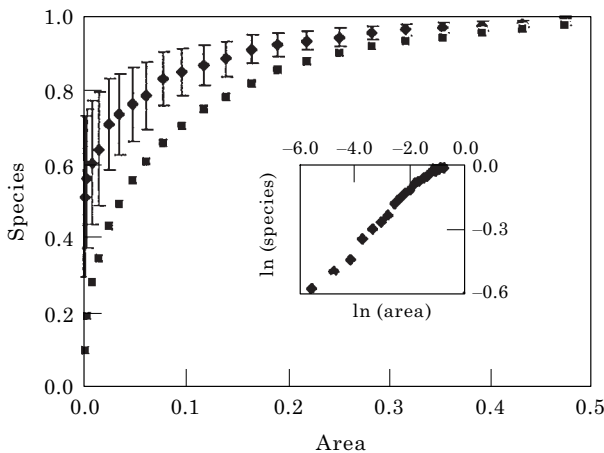


FIG. 6. The edge effect is demonstrated by the deviation between two SPARs. One is given by the average over many square patches (fluctuation between locations of the censused areas give the error bars). The lower curve is given by the number of species in a square located at one of the map's corner. It supports fewer species than the average square patch. Ranges are uniformly distributed between 0 and 100% of the total area. The SPAR can be approximated by a power law with exponent $z = 0.1 \pm 0.05$ (see \ln - \ln plot in inset). *Simulation*.

It has been suggested that the range distribution is characteristic of the size of the region (Gaston, 1994). A large region is likely to support species whose ranges are smaller than the total area available; then the distribution might look like an exponential (with few species having large ranges and many species having small ranges). If the region studied is limited to a much smaller scale, species will have ranges that overlap the entire region, and the distribution might be flat (like a uniform distribution, for instance). An intermediate size area might show a range distribution with two peaks at both extremes (large and small ranges), see Gaston (1994). If the range distribution is a truncated exponential that vanishes at about half of the total area, the SPAR is a power law [eqn (1)] with an exponent $z \approx 0.5$ [Fig. 5(b)]. By contrast, if it is a uniform distribution with minimum range 1 and maximum range equal to the size of the region, the SPAR is more concave, that is, the exponent is smaller ($z \approx 0.1$ in Fig. 6). The commonality (species in common to two patches of equal area) weakly depends on area, but decreases with distance [see eqn (30) in dimension one and Fig. 7(b) in dimension two].

Furthermore, consider two communities with increasing degree of aggregation. We vary the density within a patch, i.e. occupancy divided by range: (i) geographically restricted species occupying a single (square) patch but with a low density (0.01 in this case), (ii) restricted species occupying a single patch with density 1. This excludes cases where individuals are not distributed uniformly within a patch. Our results suggest that one can classify these two cases in one category, namely restricted species communities, since there are no significant differences in the SPAR [Fig. 5(b)]. However, going to an even smaller species density will eventually lower the SPAR, and increase the exponent z .

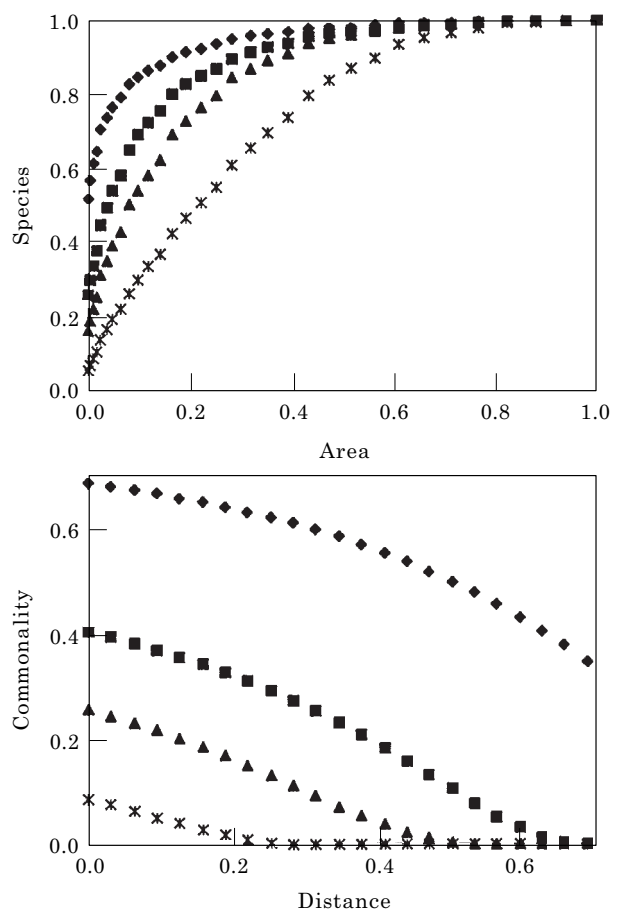


FIG. 7. SPARs (a) and commonalities (b) when ranges are uniformly distributed between 0 and $P\%$ of the total area, with $P = 100, 50, 30$ or 10 . Larger maximum ranges imply greater saturation levels. The highest curve is the same as in Fig. 6, where $z \approx 0.1$, and the lowest curve clearly yields $z \approx 1$. *Simulation*.

In Section 6.1 we compare the two kinds of communities (aggregated and widespread), and we discuss a more general community.

5. Island Species–Area Curve

To evaluate the SPAR for a collection of disconnected patches (like a set of islands), we assume that we know the range distribution, $P(r)$, and the probability that a species of range r reaches an island of area a , $p_{r,a}$. We then distribute N species on a set of islands (or patches) of various sizes $\{a\}$.

The simplest additional assumption for islands is the following: a species establishes itself on island a with probability $p_{r,a} = 1$ if its range is smaller than the island area, and $p_{r,a} = 0$ otherwise. Then, the SPAR is given by

$$S_I(a) = N \sum_r P(r) p_{r,a} = N \sum_{r \leq a} P(r) \quad (31)$$

One gets very different shapes depending on the shape of the range distribution, particularly its variance.

When variance is zero, if all ranges equal r_0 , then $S_I(a)/N$ is a step function, it is zero if $a < r_0$ and 1 otherwise [from eqn (31) with $P(r) = \delta(r_0)$]. Any distribution that is peaked at some r_0 with a small variance around this value yields a SPAR that looks like a step function but with smooth edges. If one fits a power law to this curve (or at least a large part of the curve), the power would be very small, i.e. $z \ll 1$ in eqn (1). The other limiting case is when the variance of the range distribution is very large. The maximum variance is for a uniform distribution taking values between 1 and A , the area of the largest island. Then $S_I(a)/N$ is a straight line between zero and one, that is, a power law with exponent $z = 1$. For a general distribution, with a width between 0 and A and a mean r_0 , the SPAR fits between the step function and the straight line. For a fixed mean, as the width decreases from A to zero, z decreases from 1 to 0. On the other hand, for a fixed width, the larger the mean range, the larger the area of the islands where species can establish themselves, thus the later the saturation of the SPAR (the area where the number of species reaches N).

In Section 6.2, we compare these island SPARs with nested SPARs that we obtained previously.

6. Discussion

In this section, we combine results from the previous three sections. First, we study the impact of species degree of aggregation on the SPAR and commonality curves. Second, we investigate scale effects. We finally compare SPARs for regions at various scales and also with island SPARs.

6.1. WIDESPREAD VS. RESTRICTED SPECIES

We consider a community of species that live in a single patch. Our methods generate a family of SPARs that are concave, like power laws with exponent smaller than 1 (in contrast to some logistic curves that have a point of inflection). The SPARs have distinctive shapes depending upon whether one considers a community of geographically restricted species or widespread species. In the case with restricted species, one gets a power law [eqn (1)]: the less restricted the species, the smaller the exponent with a faster saturation [Fig. 7(a)]. On the other hand, widespread species show a more abrupt change in the SPAR, which grows rapidly to saturation (Figs 1 and 2). The difference between the two types of community comes from the fact that in one case a species can be found anywhere, although it can be rare, while in the other case it is restricted locally. Therefore, we can anticipate that species (either regularly spaced or grouped in patches) distributed in the entire region, all belong to the same category of widespread species and lead to similar SPARs [see the numerical study in Section 3; and Williams (1995) for the random placement model]. Geographically restricted species are a distinctive category; in particular, our results suggested that the density within an occupied patch (the factor of proportionality between range and occupancy) is not relevant for the SPAR. The two categories share some properties, such as decreasing slope of the SPAR as the censused area increases, and they become similar if species ranges are small.

A real community will have species with various geographic ranges and occupancies. What matters then is the fraction of species with maximum degree of aggregation (restricted to one patch). From this information and the knowledge of the distribution of occupancy, one can predict the SPAR. Similarly, the SPAR and commonality reflect the spatial distribution of species, and can be used to get an idea of the fraction of restricted species and of the distribution of occupancy. For example, we simulate a community having an exponential occupancy distribution and either ranges equal to occupancy (one patch) or the whole region (uniformly distributed species). We vary the fraction of widespread vs. restricted species. The latter has to be large enough to prevent a rapid saturation of the SPAR induced by the widespread species.

The commonality computed numerically is similar to the analytical results of the previous section. The commonality has distinctive shapes for the different categories of species community. When the species are all widespread, the number of species in common to two patches does not depend on the distance between the two, but does depend on the patch areas and on the occupancy distribution [Fig. 8(a)]. In contrast, for a community of restricted species, the commonality decreases as distance increases and depends very little on the areas of the two patches considered. It vanishes when the distance is equal to the larger range linear size in the community [Fig. 8(c)]. In a community where a fraction of the species are restricted to one patch while the others are widespread, the commonality is a mixture between the two shapes and starts to decrease before it saturates at a value that depends on the patch areas [see Fig. 8(b), and Section 7 for a comparison with field data].

6.2. ISLAND VS. MAINLAND

SPARs for a collection of disconnected patches (island SPARs) are estimated in a different manner from those for a single region (mainland SPARs). To compare the two, we first specify the range distribution. For mainland SPARs, it will depend on the size of the region A considered. If A is small, then most species ranges will be large compared with the total area,

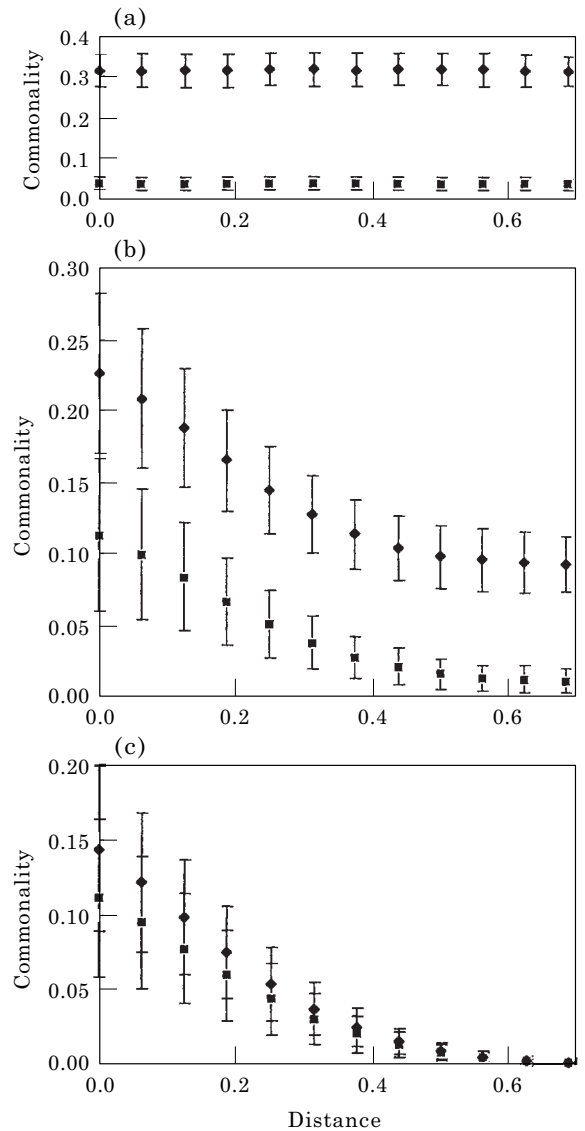


FIG. 8. Commonalities when two kinds of species are mixed: some species are uniformly distributed with occupancy of 5% of the total area, other species are restricted to one patch with range distribution as in Fig. 5(b). (a) The commonality is constant as a function of distance between patches, but varies with the area (upper curves: area is 1.5% and lower curves: 0.4%) when all species are of the first kind. (c) It decreases and vanishes at about half the largest distance when all species are of the second kind. (b) A mixture of 30% of the first kind and 70% of the second is a compromise between the two previous curves. Error bars reflect spatial fluctuations. *Simulation*.

that is, the range distribution has a large mean (in the extreme case, it is a widespread community where all ranges equal the entire area). If A is large, then most species ranges will be much smaller than A . The range distribution will have mean and variance much smaller than

A. We have already seen that larger ranges imply earlier saturation of the SPAR (the smaller the area where we detect all species), and thus smaller exponents in eqn (1), see e.g. Fig. 7(a). Therefore, different range distributions explain why exponents are smaller for small *A* SPARs (i.e. large ranges compared with *A*) and larger for large *A* SPARs. For instance, Rosenzweig's review (Fig. 9.11 in Rosenzweig, 1995) gives as typical values, $z = 0.15$, for SPAR within provinces (small *A*) and, $z = 0.9$ for SPAR between provinces (large *A*).

We do not know the appropriate range distribution that verifies our assumption that a species cannot establish itself on an island of area smaller than its range. However, consider a range distribution with narrow variance around a small mean range. We showed in Section 5 that $z < 1$. On the other hand, since small ranges, which correspond most likely to a large *A* (mainland), mean a late saturation of the SPAR and a large z , the island exponent is smaller than the mainland exponent. Next, consider range distributions with large variance. Then the opposite inequality between exponents holds: the island SPAR gives an exponent close to 1 while the mainland SPAR gives a smaller exponent. Therefore, we get that the island SPAR exponent is larger than the small *A* mainland SPAR but, smaller than the large *A* mainland SPAR. Rosenzweig gives typical values for island SPARs, $z = 0.25\text{--}0.45$ (Fig. 9.11 in Rosenzweig, 1995). If one compares with the values of the previous paragraph, one sees that the inequalities between the three exponents are in agreement with our argument (see also Rosenzweig, 1997 for a discussion on these issues).

7. Conclusions and Perspectives

We developed methods for modeling the SPAR that includes the generic properties of species of geographic range and occupancy. Our numerical methods can be used in one, two or three dimensions. They rely on three assumptions that could be released easily if field data would suggest otherwise: (1) species are distributed independently, (2) are uniformly distributed within their ranges, and (3) ranges have isotropic shapes. We need not specify the

abundance of the species (number of individuals in a cell), only their presence or absence, which is easier to observe in the field.

We generated nested SPAR for communities of species living in an area consisting of a single region. When the SPAR can be approximated by a power law, the exponent depends on the relative importance of the range distribution characteristics (mean and variance) and the size of the region. For instance, we studied a community where each species is geographically restricted, that is, occupies one patch of range given by a random distribution. We showed that larger ranges compared with the region size imply smaller exponents. If some species in the community are widespread, then the SPAR tends to be more concave than previously (as the exponent decreases). Furthermore, SPARs for communities of species living in several disconnected patches (like sets of islands) lead to exponents that are intermediate between large-scale and small-scale nested SPARs. These results are in agreement with reported field data (see e.g. Rosenzweig, 1995).

We showed that the number of species in a patch depends not only on its area, but also its shape and location, in particular we studied an edge effect that implies a diminution of the number of species near the edges. This has important implications for conservation. Taking into account spatial arrangement of species in the way we did in this paper, leads to a new SPAR that gives the number of species left after habitat loss.

Commonality curves (number of species in common to two patches as a function of their area and distance) decrease as the distance between patches increases, and vanish when one reaches the larger range linear size of the community. In the limiting case when species are widespread, the commonality depends solely on area. Such behavior has been observed in field experiments: for plants in a meadow in Colorado (Harte *et al.*, 1997b), species in a forest in Carolina (Palmer & White, 1994), trees in a mountain in Oregon (Whittaker, 1972), and birds in Australia (Cody, 1993). Commonality is not used consistently in the literature. For example, Harte *et al.* (1997a, b) show a plot of our commonality divided by the number of

species in each of the two patches considered; it decreases with distance but shows no saturation. This probably means that some species have a restricted range (implying distance dependent commonality) while other species can be found everywhere in this region (since the commonality does not saturate as a function of distance, and it depends on area). Palmer & White (1994) show a plot of the number of species in each of the two areas considered minus the commonality (Fig. 4 in Palmer & White, 1994). Their curves are either flat or increasing with distance (cf. our Fig. 8). A more systematic study of field data is needed to check that what we predict for the range distribution by looking at the species–area and commonality–distance curves is actually verified.

Finally, our work suggests a number of potential avenues of future research:

- more work needs to be done to confront models of SPARs to field data;
- careful empirical attention to our method and others by which SPARs are created and the different methods should be compared at the same site;
- solution of the inverse problem relating commonality and SPAR to the spatial distribution of the species;
- a dynamical model might be coupled to our model: starting from the map of species that we were able to create, one could study the evolution and stability of the spatial patterns when dispersal, life cycles and competition are taking into account (e.g. Durrett & Levin, 1996). This could be useful especially when studying species displacement after habitat loss.

The work of M. Ney-Nifle is supported by the Centre National de la Recherche Scientifique (CNRS-France). We thank Karen Holl for introductions. We are very grateful to John Harte and Mike Rosenzweig for sharing their papers before publication, and for useful comments on a previous version of this paper.

REFERENCES

- BUCKLEY, R. (1982). The habitat-unit model of island biogeography. *J. Biogeography* **9**, 339.
- CODY, M. L. (1993). Bird diversity components within and between habitats in Australia. In: *Species Diversity in Ecological Communities* p. 147. Chicago, IL: University Chicago Press.
- COLEMAN, B. D. (1981). On random placement and species–area relationship. *Math. Biosci.* **54**, 191–215.
- COLEMAN, B. D., MARES, M. A., WILLIG, M. R. & HSIEH, Y.-H. (1982). Randomness, area, and species richness. *Ecology* **63**, 1121–1133.
- COLWELL, R. K. & HURTT, G. C. (1994). Nonbiological gradients in species richness and a spurious rapoport effect. *Am. Nat.* **144**, 570–595.
- CONNOR, E. F. & MCCOY, E. D. (1979). The statistics and biology of the species–area relationship. *Am. Nat.* **113**, 791–833.
- CRESSIE, N. (1991). *Statistics for Spatial Data* p. 570. New York: Wiley-Interscience.
- DURRETT, R. & LEVIN, S. (1996). Spatial models for the species–area curve. *J. theor. Biol.* **179**, 119–127.
- FELLER, W. (1950). *An Introduction to Probability Theory and its Applications* Vol. 1. New York: Wiley.
- GASTON, K. J. (1990). Patterns in the geographical ranges of species. *Biol. Rev.* **65**, 105, and references therein.
- GASTON, K. J. (1994). *Rarity*. London: Chapman & Hall.
- GOTELLI, N. J. & GRAVES, G. R. (1996). *Null Models in Ecology*. Washington DC: Smithsonian Institution Press, see Species–area relationships and Biogeography chapters.
- HARTE, J. & KINZIG, A. P. (1997a). On the implication of species–area relationship for endemism, spatial turnover, and food web patterns. *Oikos* **80**, 417–427.
- HARTE, J., MCCARTHY, S., TAYLOR, K., KINZIG, A. & FISHER, M. L. (1997b). Estimating species–area relationship from plot to landscape scale (submitted for publication).
- HILBORN, R. & MANGEL, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton, NJ: Princeton University Press.
- HOLLAND, H. D. & PETERSEN, U. (1995). *Living Dangerously. The Earth, its Resources, and the Environment*. Princeton, NJ: Princeton University Press.
- KAWASAKI, T. (1979). Time evolution of a magnetic binary alloy. *Prog. Theor. Phys.* **61**, 384–400.
- KRAUTH, W. (1996). Introduction to Monte Carlo algorithm (text and codes at the author's home page <http://www.lps.ens.fr/~krauth/intromonte.html>).
- LEGENDRE, P. & FORTIN, M.-J. (1989). Spatial pattern and ecological analysis. *Vegetatio* **80**, 107–138.
- LEITNER, W. A. & ROSENZWEIG, M. L. (1997). Nested species–area curves and stochastic sampling: a new theory. *Oikos* **79**, 503–512.
- MACARTHUR, R. & WILSON, E. O. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.
- MCGUINNESS, K. A. (1984). Equations and explanations in the study of species–area curves. *Bio. Rev.* **59**, 423–440.
- MAUER, B. A. (1994). *Geographical Population Analysis: Tools for the Analysis of Biodiversity*. Oxford: Blackwell.
- PALMER, M. W. & WHITE, P. S. (1994). Scale dependence and the species–area relationship. *Am. Nat.* **144**, 717–740.
- PRESTON, F. W. (1962). The canonical distribution of commonness and rarity. *Ecology* **43**, 185–215, 410–432.
- ROSENZWEIG, M. (1995). *Species Diversity in Space and Time*. New York: Cambridge University Press.
- ROSENZWEIG, M. (1997). Species diversity. In: *Theoretical Ecology* (MacGlade, J., ed.) 3rd Edn. Oxford: Blackwell.

- ROSENZWEIG, M. (1998). Preston's ergodic conjecture: the accumulation of species in space and time. In: *Biodiversity Dynamics: Turnover of Populations, Taxa and Communities*. (McKinney, M. L., ed.). New York: Columbia University Press.
- SIMBERLOFF, D. (1992). Do species-area curves predict extinction in fragmented forest? In: *Tropical Deforestation and Species Extinction*. London: Chapman & Hall.
- WHITTAKER, R. H. (1972). Evolution and measurement of species diversity. *Taxon* **21**, 213–251, see also Pielou, E. C., *Ecological Diversity*. New York: Wiley-Interscience.
- WILLIAMS, M. (1995). An extreme-value function model of the species incidence and species-area relations. *Ecology* **76**, 2607–2616.

APPENDIX

Generating Patchy Spatial Distribution of Species

A.1. GENERAL PICTURE AND EXAMPLES

To generate the spatial distribution of species, we need an algorithm that applies for widespread, patchy or localized species arrangements. We use the Kawasaki model, introduced in statistical physics to study atomic aggregation (Kawasaki, 1979). One starts with a random configuration of individuals of a certain kind of species and begins with a map divided into cells that are black (white) for occupied (empty) cells; the black and white cells being mixed in a way such that one can not distinguish any pattern. If this species is meant to be an aggregated species, then one has to group the black cells in patches, the more aggregated the smaller the number of patches. The Kawasaki algorithm does that by introducing an attractive interaction between the occupied cells. As the simulation time increases, species become more and more aggregated, i.e. time plays the role of the degree of aggregation.

In Fig. 4(a,b), we show the evolution of two configurations that have different densities of black cells. This density is the occupancy of the species under consideration whereas time can be translated into an average patch area (related to the geographic range since the longer the simulation goes the more aggregated the species, i.e. the smaller its range). This correspondence can be done empirically or using the fact that the

average patch area grows with time as a power law (for a range of values of K , see Kawasaki, 1979).

A.2. THE ALGORITHM

The Kawasaki algorithm works as follows. One assigns an "energy" to each configuration of black and white cells in a way so that the preferred configuration is a segregation of the colors, e.g. a single black patch in a white background. If one starts with a random configuration, the dynamics lead the system towards its most energetically favorable configuration. The model fulfills this requirement by giving a rule (described below) that generates a sequence of configurations that is a Markov chain (Cressie, 1991). We generate a new configuration from the previous one by Monte Carlo methods (Krauth, 1996; Hilborn & Mangel, 1997).

We start from an initial configuration I (a random distribution of black cells, in our case). The next configuration in the Markov chain, F , is obtained by considering a pair of neighboring cells and exchanging their color if they happen to be different. This can be seen as the move of one color until it reaches a patch of the same color. The difference in energy between the two configurations is then determined using the Ising model (Cressie, 1991; Krauth, 1996).

$$E(C) = -K \sum_{\langle i,j \rangle} \sigma_i^C \sigma_j^C \quad (\text{A.1})$$

where σ_i^C takes the value 1 if the cell is black and -1 if it is white. The sum is over neighboring cells (each cell has four neighbors, except on the edges). To minimize this energy one has to keep the product $\sigma_i^C \sigma_j^C$ positive, that is, neighboring cells prefer to be of the same color. In other words, the system will tend to form patches that will grow through time as cells of the same color aggregate. The parameter K controls the shape and the rate of formation of the patches. Here, we fixed $K = 0.25$, which is larger than a critical value below which the system would tend to equilibrium configurations with no spatial pattern (Kawasaki, 1979).

Finally, we compute the probability of moving to the configuration F [using the Metropolis transition probability (Cressie, 1991; Krauth, 1996)]

$$p(I \rightarrow F) = \min[1, \exp(E(I) - E(F))] \quad (\text{A.2})$$

This is implemented using a single uniformly

distributed random number $0 < U < 1$, and we move to the new configuration under the condition that $U < p(I \rightarrow F)$. Therefore, the transition probability is 1 if the final configuration energy, $E(F)$, is smaller than the initial configuration energy, $E(I)$.