Evolutionary Ecology Research, 2003, **5**: 469–492

A unified theory for macroecology based on spatial patterns of abundance

Brian McGill[1]* and Cathy Collins[2]

[1]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721 and* [2]*Yosemite National Park, PO Box 907, CA 95389, USA*

ABSTRACT

Macroecology proceeds by identifying patterns and then identifying processes that cause those patterns. Most of the processes that macroecologists study are local in nature and tend to involve species interactions and speciation and extinction processes. In contrast, we propose that several important macroecological patterns can be explained by very large-scale processes that are primarily spatial in nature. Specifically, we suggest that the structure of abundance across a species' entire range combined with interspecific patterns in range location and global abundance can explain the well-known macroecological patterns of: (1) a positive correlation between range size and abundance, (2) the species–area relationship, (3) decay of species similarity with distance and (4) the species abundance distribution. We show that spatial processes produce these patterns through a combination of analytical and Monte Carlo analysis. We also show that the connection is robust (indifferent) to the precise mathematical assumptions. Such a theory might be called a unified theory, because it explains multiple patterns with a few processes. To differentiate among the growing number of unified theories, we suggest that testing additional predictions over and above producing curves of the correct shape is important. To this end, we present several novel, quantitative predictions and provide empirical tests. In short, we provide an empirically grounded and tested theory, which suggests that superimposing individual species ranges across space creates local community patterns.

Keywords: macroecology, species abundance distribution, species–area relationship, species range, unified theory.

INTRODUCTION

Macroecology is the study of large-scale patterns and processes (Brown, 1995). One of the key goals in macroecology is to identify the processes that underlie the patterns that are found. Because of the difficulty of conducting controlled, replicated experiments at these scales, theory plays an important role in identifying these processes. Recently, researchers (Hanski and Gyllenberg, 1997; Harte *et al.*, 1999; Gaston and Blackburn, 2000; Hubbell, 2001) have gone one step further and noted that theory can be used to tie together several separate patterns that can be explained by a common, small set of processes. This is a

* Author to whom all correspondence should be addressed. e-mail: mail@brianmcgill.org

'unified' theory, whether it be in physics or macroecology. Unified theories appeal strongly to our sense of elegance and parsimony.

It should be noted that unified theories in macroecology are not new. Both Preston (1962) and May (1975) noted a connection between two of the most important patterns in macroecology, the species abundance distribution (hereafter SAD) and the species–area relationship (hereafter SPAR). Indeed, mathematically, there must be a connection. If we know the distribution of abundances, then we know what the sampling distribution should look like. In particular, we know what the collector's curve that plots number of species sampled versus number of individuals sampled will look like. At very small scales and assuming density of individuals per unit area is constant, then the species–area curve is really nothing more than a collector's curve (Preston, 1962; Rosenzweig, 1995; Hubbell, 2001). Hence, there is a tight mathematical link, at least at some spatial scales.

Given the fact that unified theories of macroecology are not novel and are now becoming common, we must ask how we choose among the various theories proffered. Part of the answer is that we may not have to completely choose – more than one mechanism can be in operation and usually is in ecology. Furthermore, different unified theories may apply at different spatial scales. Thus, it is important that we start attaching explicit spatial scales to these theories. But we should not use this as an excuse to accept all unified theories offered. One important criterion that we must examine is testability. A scientific theory is testable to the extent that it makes predictions. The extent to which the predictions are new (i.e. not explaining data that is already known) is important. So is the extent to which the predictions are falsifiable. A prediction of a general curve shape where the parameters must be chosen to maximize the fit to the data is much weaker than the prediction of a curve shape where the parameters are derived from the theory. Similarly, a theory that makes a single qualitative prediction is weaker than a theory that makes multiple qualitative predictions and/or precise quantitative predictions.

In this paper, we present a unified theory of macroecology based on spatial structure of abundance. Specifically, it relies on patterns at the very large spatial scales of entire species ranges – that is, the scale of significant portions of continents. This contrasts with many prior macroecological theories, which have relied on local processes such as population dynamics and species interactions. Despite the large-scale patterns upon which our theory is based, it makes predictions about relative abundance (SAD) at the scale of a community, an area containing species that interact with each other and which is imprecise but on an order of magnitude of hectares or square kilometres. This theory suggests a common set of processes behind four well-known macroecological patterns. In the spirit of the previous discussion, it is our intention to go further and make a strongly testable unified theory. We will present three additional novel, more quantitative predictions and then test these using empirical data on North American birds.

PREMISES

We start with three empirically well-documented facts:

Premise 1: Each species' geographic range is located in space relatively independently of the others

Four empirical studies have found that the modes of species abundance (and also range boundaries when tested) are more often than not statistically indistinguishable from a

random placement (Austin, 1987; Shipley and Keddy, 1987; Minchin, 1989; Hoagland and Collins, 1997). For those taxa which are not random, they are usually clustered (under-dispersed) (Austin, 1987; Shipley and Keddy, 1987; Minchin, 1989; Hoagland and Collins, 1997). *Note*: the fact that the placement fits a random model in no way implies that the location of species is not a deterministic process based on climate and species interactions (Pielou, 1977). Rather, it suggests that so many deterministic factors are involved that statistical laws apply and that *pairwise* species interactions do not dominate.

Premise 2: Different species vary in global abundance according to a hollow curve

In this paper, we define global abundance to be the highest observed abundance at any single location across the entire range. Results are very similar if we use total abundance (all individuals in the range) or average abundance (averaged at each site where the species are found). Each of these measures of global abundance demonstrates an interspecific hollow curve distribution of abundances (i.e. a histogram gives a monotonically decreasing function with apparent asymptotes at the y- and x-axes). This pattern is usually observed to describe abundances in a local community, but it has also been recognized to describe the distribution of abundances at a global scale – for example, Fig. 1 and results presented later (see also references in Gaston, 1994, pp. 33–34; Gaston, 1996a).

Premise 3: Abundance across a range is structured according to a 'peak-and-tail pattern'

We use the phrase 'peak-and-tail pattern' to describe a pattern of abundances across a range that has a few (1–5) peaks of very high abundance which drop off rapidly but fairly smoothly to a long tail which is 1–2 orders of magnitude lower in abundance than the peaks. We know that species abundances vary significantly across their range (Darwin, 1859; Grinnell, 1922; Brown *et al.*, 1995), being rare more often than common (Brown, 1995; Murray *et al.*, 1999). Although a symmetric pattern with the peak in the centre (i.e. Gaussian) is not uncommon (Austin, 1987; Brown, 1995; Brown *et al.*, 1995, 1996), it is also common to find asymmetric, non-centred and even multi-peaked distributions (Austin, 1987; Minchin, 1989; Sagarin and Gaines, 2002). The peak-and-tail description includes these cases, while the stronger statement that abundances are normal or Gaussian across their range (Whittaker, 1951) does not.

Hereafter, we call these three facts 'premises' of our model. We choose the term 'premise' carefully. It is stronger than the term 'assumption' – there is strong empirical evidence for them. At the same time, we recognize that the three premises do not represent ultimate mechanisms and hence avoid calling them mechanisms; there clearly must be a more mechanistic explanation behind our three premises.

MATHEMATICAL MODEL

We now translate these three premises into the language of mathematics. As always in mathematical models, there is a trade-off between analytical tractability and accuracy. To address this, in each case we will perform robustness analysis to show that the loss of reality caused by choosing simpler, more mathematically tractable descriptions does not affect our results. Thus the three premises in mathematical terms are:

1. *The location of the peak of each species range, μ_i, is distributed according to a Poisson spatial process* (Taylor and Karlin, 1998). Since there is considerable evidence that

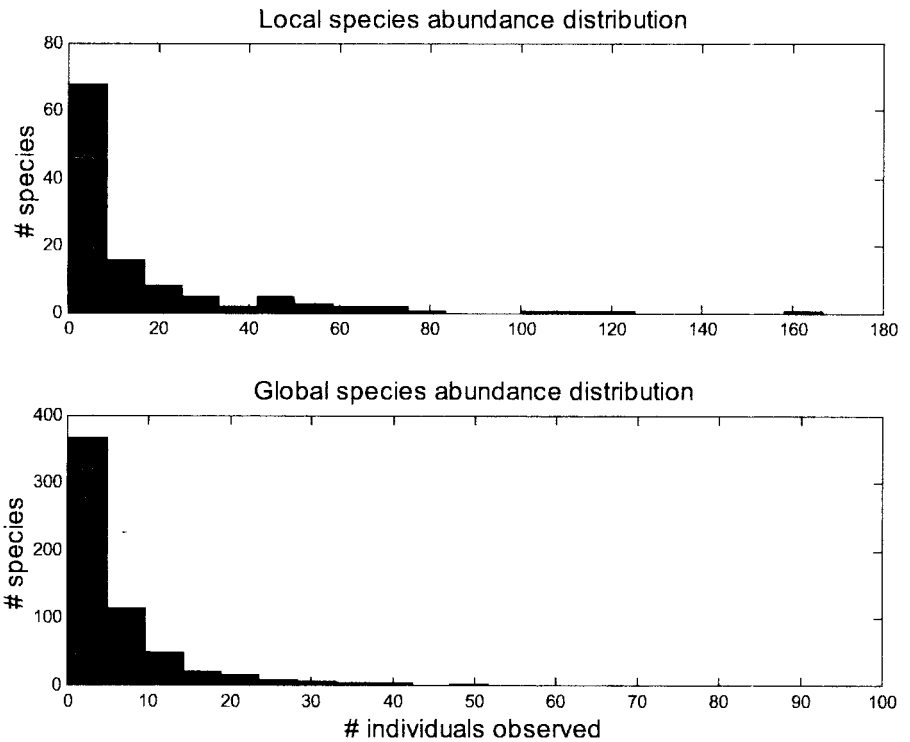

Fig. 1. The top graph is an empirically observed example of a local species abundance distribution (SAD) for the BBS route at 100.217°W latitude, 48.883°N longitude in North Dakota near the Canadian border. There are many species with a very low abundance (left side of graph). There are a few species with a high abundance (right side of graph). This shape is sometimes called a 'hollow curve'. The bottom graph gives a species abundance distribution for global abundance. In this case, global abundance is measured by abundance averaged across the range (AVGABUND), but the pattern is similar for other measures of global abundance. Note that the global SAD also is distributed with a hollow curve, just as local SADs are. Our model predicts that these two curves should have the same shape (although not the same scale). We examine quantitative evidence for this similarity in the text.

ranges of a few (but not most) taxa are clumped, we also perform Monte Carlo simulations using a Neymann-Scott model (Stoyan and Stoyan, 1994). The clumping turns out to make very little difference.

2. *The peak abundance, $NMAX_i$, of any species is distributed according to a log-normal distribution*. Other distributions probably fit better, but there is to date no consensus on what is the correct distribution. In fact, rarely if ever has anyone tried to fit a distribution other than the log-normal to a distribution of global abundance (as opposed to the more commonly studied local abundance distributions, which has dozens of proposed distributions). There is demonstrably a good fit of the log-normal to the data (shown later in the paper). As a form of sensitivity analysis, we will also use the power distribution ($p(x) = k(x/b)^c$) (Evans *et al.*, 1993). The power distribution also possesses a hollow curve shape. Although it fits our data less well than the log-normal, it is highly tractable

analytically. The fact that all of our Monte Carlo simulations perform very similarly whether we use the log-normal or the power distribution suggests that our model is robust to the exact shape of the distribution and especially to differences in distribution on a log-scale. This suggests that the only requirement for our model to work is that a hollow-curved distribution be used.

3. *The 'peak-and-tail' structure of abundance does not have an obvious translation into simple mathematics*. Some authors have claimed that a Gaussian (bell) curve over space represents the abundance of each species over its range (Gauch and Whittaker, 1972; Austin, 1987; Brown, 1995). Others have argued for asymmetric and/or multi-peaked shapes (Austin, 1987; Minchin, 1989; Sagarin and Gaines, 2002). For our analytical models, we use a Gaussian curve. However, we also perform sensitivity analysis to determine the dependence of our models on the Gaussian shape. In one case, we check our results with Monte Carlo simulations using a sum of Gumbel extreme value distributions (Evans *et al.*, 1993) to represent a multi-modal, asymmetrical distribution of abundances. This much more complicated distribution provides very similar results. In another case, we can decouple our model entirely from the exact structure of abundance within a range, as all we need to know is the distribution of range sizes. Details of the robustness analyses are summarized in Table 1. Examining the inner workings of the models suggest that all we really require is: (1) high, small peaks; (2) large, low tails; and (3) a relatively smooth transition between the two. In parts of our model, we represent space as one-dimensional and in other cases as two-dimensional.

Figure 2 pictorially presents these premises as used in the analytical models. Thus the abundance at any point in space, say X, for the ith species is given by:

$$N_i(X) = \mathrm{NMAX}_i \exp\left(-\frac{\|X-\mu_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

where σ_i is a scaling constant for range size (representing the distance from the centre, μ_i, to the inflection point where abundance drops off most rapidly).

We need to add one additional, technical assumption. This is because Gaussian curves have infinite spatial extent (their tails go to infinity). To fix this, we assume that all species ranges terminate (reach their edge) when the abundance drops to some fixed cut-off, NMIN_i. This is the same criterion used to delimit range boundaries in the Breeding Bird Survey (BBS) atlas (Price, 1970). Hence, the range boundary occurs on the circle $\{X: N_i(X) = \mathrm{NMIN}_i\}$.

FOUR QUALITATIVE PATTERNS

These three premises alone (plus the technical assumption) are enough to produce four well-known and, in many cases, poorly explained macroecological patterns:

1. The positive correlation of range size with abundance
2. Species–area relationships (SPARs)
3. Species turnover (similarity decay) patterns
4. Species abundance distributions (SADs)

In short, *superimposing individual species ranges across space creates local community patterns*. We will now briefly describe each of the four patterns and demonstrate through

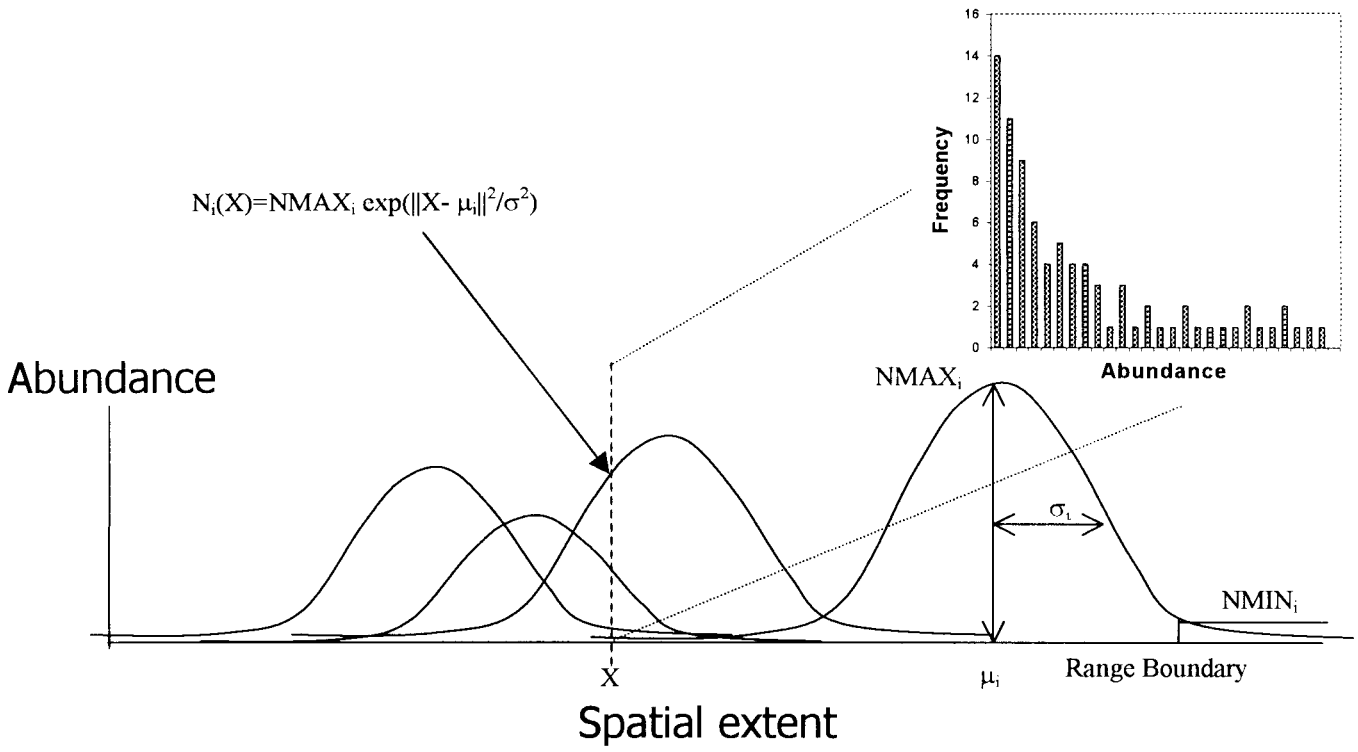

Fig. 2. A pictorial representation of the model. Each species has a range that is placed independently in space with a centre at μ_i. The peak height, $NMAX_i$, varies by species as does the width, σ_i. Over a species range, the abundance varies according to the formula for a Gaussian curve. The range boundary occurs where the abundance drops to $NMIN_i$. This picture also shows how this model generates SADs. If we sample at any one point, we get some species with high abundance that are near the centre of their range and some species that are at low abundance because they are near the edge of their range. As we add additional species, we build a SAD graph that appears like the one shown in the pullout.

Table 1. A summary of the robustness tests performed

Mathematical premises	Range size – abundance correlation alternative	SAD alternative	SPAR alternative
1. Poisson distribution of range centres	N/A	Neymann-Scott model (Stoyan and Stoyan, 1994) of a clustered random process ($\lambda = 5$ and points placed exponentially far away in an isotropic fashion) which was checked using V/M and A measures of clustering (Pielou, 1977) to verify that it produced statistically significantly clumped points	Same as SAD
2. Log-normal distribution of global abundances	N/A	Power distribution with shape exponent of 0.298 (the MLE estimate using the BBS AVGABUND)	Same as SAD
3. Gaussian shape of abundance across a range	Derived formula – equation (2) – is recognized as only approximate	Sum of three Gumbel extreme value distributions with peak of each distribution chosen randomly (multi-peaked with each peak asymmetric)	Premise used only to produce RADIUS_i. Thus the resampling method of calculating RADIUS_i provides an alternative test
4. Minimum population size	Exact formula (equation 2) is incorrect, but more general qualitative model described in text still holds	Analytical model has no lower limit. Monte Carlo simulations performed with and without limit	As above

Note: The robustness tests demonstrate that the simplifications made when translating the empirically observed premises into mathematical language are not vital to the results of the models. In short, the models are robust to fairly significant deviations of the mathematical details. *The alternative mathematical assumptions described above had no noticeable impact on the model results.*

analytical and simulation models that our three premises generate them. A summary of the logic of which premises produce which patterns can be found in Fig. 3.

Pattern 1

Many authors (Hanski, 1982; Brown, 1984; Gaston, 1996b) have noticed that the range size of a species increases with various measures of abundance. Our model predicts this correlation. Solving equation (1) algebraically for the radius at which $N_i = \text{NMIN}$ and assuming circular ranges so that $\text{RangeSize} = \pi r^2 = \pi x^2$, where x is chosen such that $N_i(x) = \text{NMIN}$, gives:

$$\text{RangeSize} = 2\pi\sigma_i^2 \ln(\text{NMAX}_i/\text{NMIN}) \quad (2)$$

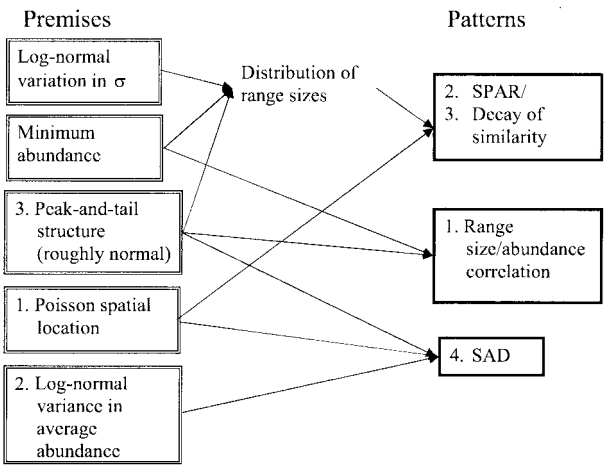

Fig. 3. An idea map of which premises and assumptions lead to which resultant patterns. The three premises are shown on the left, as are two technical assumptions. An arrow indicates that the assumption/premise is necessary to produce the pattern pointed to. For example, producing the SAD pattern depends only on premises 1, 2 and 3. The SPAR and decay of similarity patterns depend on the two assumptions and premise 3 only through the distribution of range sizes.

If we assume σ and NMIN are constant across species, with NMAX varying (as per Premise 2), then we get a monotonically increasing relationship that will produce a correlation. We do not wish to place too much emphasis on the precise nature of formula (2) for two reasons. First, the exact nature depends heavily on the assumption of a Gaussian curve, something that is only approximately correct. Rather, we see this formula as a mathematically simple example of a more general phenomenon – if abundance falls off reasonably smoothly from a peak abundance to a fixed minimum abundance (NMIN), then range size must be correlated with the peak abundance. Second, real-world examples of this correlation contain far too much scatter to be able to test the exact nature of the relationship (e.g. linear in log–log *vs* semilog). To this end, we note that if we allow σ and/or NMIN to vary by species, we will still get a positive correlation, but the scatter approximates that found in the real world (we performed this in a very simple Monte Carlo simulation where σ varied according to a power distribution fit to the BBS range sizes).

Pattern 2

The species–area relationship (SPAR) describes the fact that the number of species found fits a power law function of area ($S = cA^z$). The z-values range from about 0.05 to 0.3 (if the areas are sampled from within a single biogeographical province) (Rosenzweig, 1995). Ecologists have known about SPARs longer and documented them more thoroughly than any other macroecological pattern (Rosenzweig, 1995). Our simulations implement the three premises and the technical assumption in a spatially explicit simulation (see Appendix 1 and Fig. A1 for details). Each simulation then draws progressively larger boxes around a starting point and counts a species as present if the box intersects any part of its range. This allows us to plot a nested species–area curve (see Fig. 4).

These simulations produce a power law SPAR. Over a very wide range of parameter values, the SPARs have an average R^2 of 0.91 for a linear regression in log–log space (i.e. a

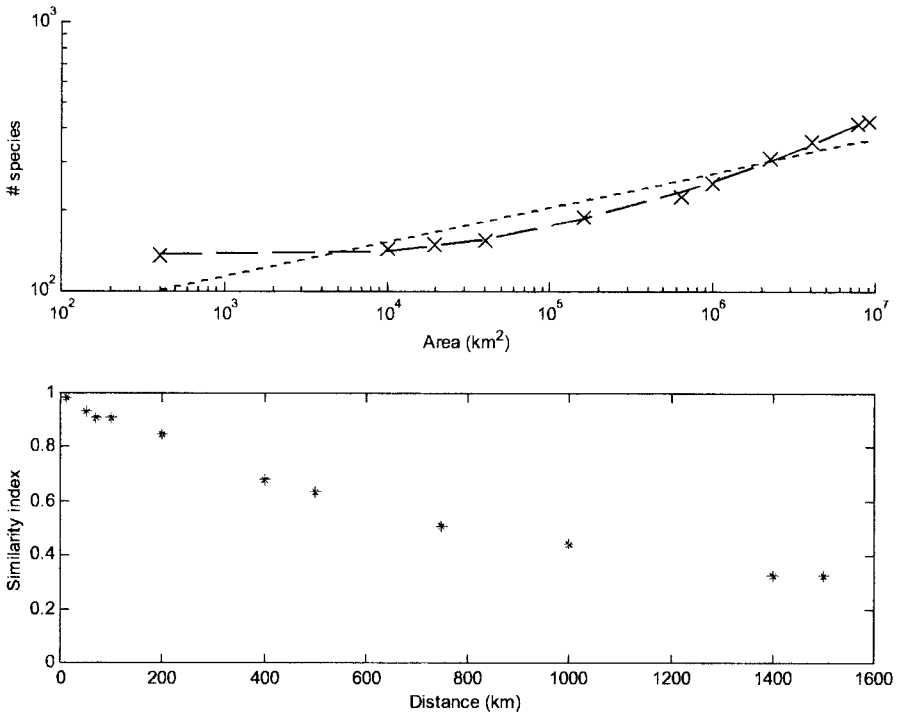

Fig. 4. The top figure shows a classic species–area relationship (SPAR) with log area (in km^2) on the x-axis and log number of species on the y-axis. The × symbols represent the simulation based on the theory in this paper with parameters as specified in Appendix 1 and the resampling method used. The dotted line is a straight-line (power law) fit to the simulation, whereas the dashed line is a quadratic fit to the simulation. The bottom plot shows a similarity–decay relationship. The y-axis plots similarity of species composition between an arbitrary starting point and various other points along a transect from the starting point. In this plot, community similarity is calculated using the Jaccard index. The x-axis plots linear distance in kilometres between the arbitrary starting point (here the centre of the square in the simulation) and select other points along a transect from the starting point (here chosen to be points due east of the starting point at various distances).

power law in arithmetic space), and the z-values range from about 0.05 to 0.25. The curves also exhibit an acceleration in slope (concave upwards) at very large areas, a pattern that is found repeatedly in nature (Rosenzweig, 1995; Hubbell, 2001). This model produces completely horizontal SPARs ($z = 0$) at very small spatial scales because we ignore the sampling effects and habitat heterogeneity that drive SPARs at this scale (Rosenzweig, 1995). Again, robustness analysis showed that the model was insensitive to the exact mathematical assumptions (Table 1).

Pattern 3

Another well-documented pattern is the increasing decay of similarity in species composition with increases in geographic distance. Ecologists call this 'species turnover', 'decay of similarity' or 'β-diversity' (in a slightly different context). When similarity is plotted

against distance, we expect to see a roughly hyperbolic curve (rapid decline over short scales, a slower decline over a long range of intermediate scales, and then an even slower decline over very long scales) (e.g. Condit *et al.*, 2002). We can use the same model that we used for SPARs to simulate species turnover. Using the Jaccard measure of similarity between points at different distances, we can plot a species similarity versus distance curve (see Fig. 4). We did not fit the data to a particular curve, as there is no well-accepted, theoretically expected curve. However, each simulation run is visually similar to those found in nature (i.e. hyperbolic).

Pattern 4

Species abundance distributions (SADs) have puzzled ecologists for nearly a century. The pattern is clear: in a given community, most species are rare and a few species are abundant (Darwin, 1859; Preston, 1948; MacArthur, 1972). Plotting the number of species against the number of individuals in each species (as in a probability distribution) yields a characteristic hollow curve called a species abundance distribution (see Fig. 1). This same curve has been plotted on a variety of different axes: log abundance on the x-axis sometimes but not always gives a modal (humped) pattern (e.g. Preston, 1948), percent or percentile on the x-axis gives a U-shaped distribution as the long, right tail bends up on the right (Raunkiaer, 1934; Hanski, 1982), while some authors plot abundance on the y-axis and numerical rank (i.e. 1st, 2nd, 3rd, etc.) on the x-axis (Whittaker, 1965), but all are mathematically equivalent.

It is easy to see how the three spatial premises we identify could lead to a SAD. If the spatial abundance distribution across a single species range has a 'peak-and-tail' structure, and if the centre of different species ranges are independently distributed, then sampling at any point in space yields an assemblage containing many rare and few common species (Fig. 2). This works because the tails are long – that is, the species are present in low abundance at most sites and common at only a few. For an empirical example of how the rank of a species within a community (i.e. the SAD) is driven by our premises, see Fig. 5. A qualitatively similar idea was suggested by Enquist *et al.* (1995) and by Gauch and Whittaker (1972).

This conceptual model can be made more rigorous by an analytical model which shows that our three premises are by themselves sufficient to generate SAD curves like those seen in nature. We use the same mathematical notation as before for simplicity (again see Fig. 2 and equation 1) and use a one-dimensional space (the interval [0,1]). Assuming a power distribution of global abundances and with some algebra (see Appendix 2), we see that the probability density function (pdf) for abundances of species at one location has the hollow curve shape characteristic of SADs. In particular, it has a power distribution with the same shape parameter as the distribution of global abundance. We will test this shortly.

We performed a series of Monte Carlo simulations to test the effects of perturbation from the mathematical assumptions used in the analytical model. Table 1 summarizes tests performed to demonstrate that our model is robust – that is, that the exact mathematical translations that we identified above are not critical to our results. In addition to the variations described in Table 1, we also performed Monte Carlo analysis with relatively low (50) versus high (5000) species diversity and with range sizes (σ) constant or variable between species. We found that none of these alternatives (individually or collectively) modifies the results – we always get a hollow curve (well fit statistically by a power distribution).

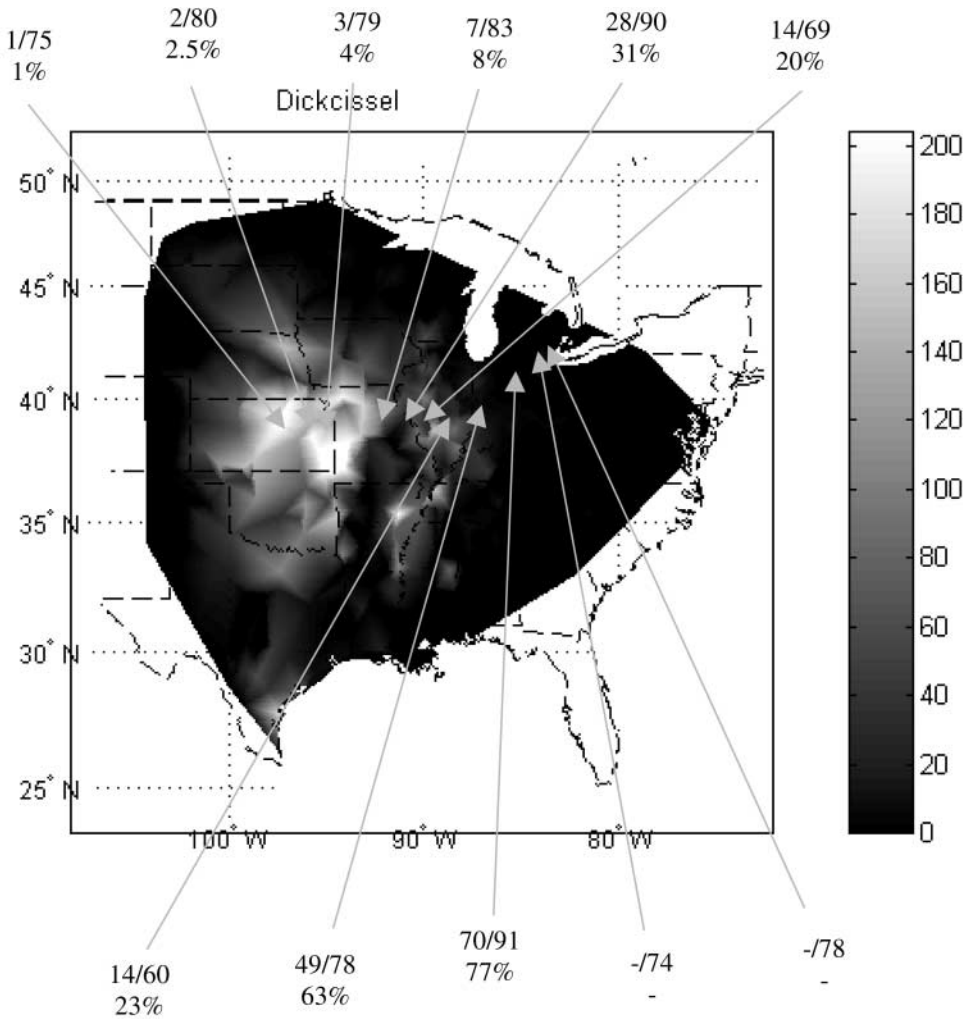

Fig. 5. A plot of variation in abundance of the Dickcissel (*Spiza americana*, in the Cardinal family) over its species range. At 11 selected points, each successively further from the peak abundance, the pullouts show the ranks of the species in terms of local abundance within the community (1 is highest) relative to total number of species in the community and the percentile of abundance within the community (to adjust for communities that differ in number of species). A ' — ' indicates that the species is absent from that community even though it is within the range. Note that as we get further from the peak, the abundance and local rank drop off; in particular, they drop off quite quickly from the peak and then have a long tail of low abundance.

EMPIRICAL CONFIRMATIONS

So far, we have shown through analytical models and Monte Carlo simulations that our three premises (and the technical assumption) can produce curves whose shapes match the four macroecological patterns often observed in nature (using eyeball tests or curve-fitting

with several free parameters). We will now present a number of additional tests of our theory against empirical data.

For data, we used the North American Breeding Bird Survey (Patuxent Wildlife Research Center, 2001; Price *et al.*, 1995; Sauer *et al.*, 1997) (see Appendix 3 for details and for definition of the variables used herein). From it we obtained several results. First, we will present results that confirm one of our three premises or one of our four patterns specifically for birds of North America:

1. *Premise 2: Different measures of global abundance are highly correlated with each other* (see Table 2). The total abundance of a species across its entire range, its average abundance and its peak abundance are all highly correlated. This suggests that the concept of global abundance is robust and lessens the need to be precise about the definition.
2. *Premise 2: Interspecific distribution of global abundance has a hollow curve distribution* (see Fig. 1). Both a power and a log-normal distribution fit well ($r^2 = 0.93$ and 0.997, respectively). Maximum likelihood estimates for the parameters are $c = 0.293$ and $\mu = 3$, $\sigma = 1.3$, respectively. The hollow-curve variation in global abundances is an assumption of our analytical model and it is here shown to be true for North American birds.
3. *Pattern 1: Range size increases with a species' global abundance*. Regression of number of routes versus ln(MAXABUND) gives an $R^2 = 0.16$ with $P < 0.001$. Using an alternative measure of range size (calculated as area in km^2 of the convex hull), regression versus ln(MAXABUND) gives a significant ($P < 0.001$) relation of range size $= 1.15 \times 10^6$ km^2 $+ 872842 \times \log(\text{MAXABUND})$ with an $R^2 = 0.13$ ($n = 457$ for both regressions) (see Fig. 6). This confirms the first macroecological pattern we are trying to explain as being true in North American birds. As with other examples of this pattern, there is considerable scatter.
4. *Pattern 4: Interspecific distribution of local abundance* (*at a given site*) *fits both the power and log-normal distributions well*. When comparing the observed to the predicted cumulative distribution functions (cdf), raw R^2 and correlational R^2 (Wilkinson, 1997, p. 451) average 0.71 (0.52) and 0.90 (0.78) for the power distribution and 0.98 (0.95) and 0.98 (0.96) for the log-normal distribution, respectively (where the figures in parentheses represent the lower limits of a 95% confidence interval).

NOVEL, QUANTITATIVE PREDICTIONS TESTED

We now move from confirmatory results in the previous section and from qualitative, curve-fitting results in the section before that to a stronger test of our theory by examining novel, quantitative predictions that our model makes and by testing them with the BBS data.

Table 2. Pearson correlation coefficients (r) between various measures of global abundance

	TOTABUND	MAXABUND	AVGABUND
TOTABUND	1.0000	0.8832	0.7114
MAXABUND		1.0000	0.8735
AVGABUND			1.0000

Note: All correlations are significant at $P < 0.0001$.

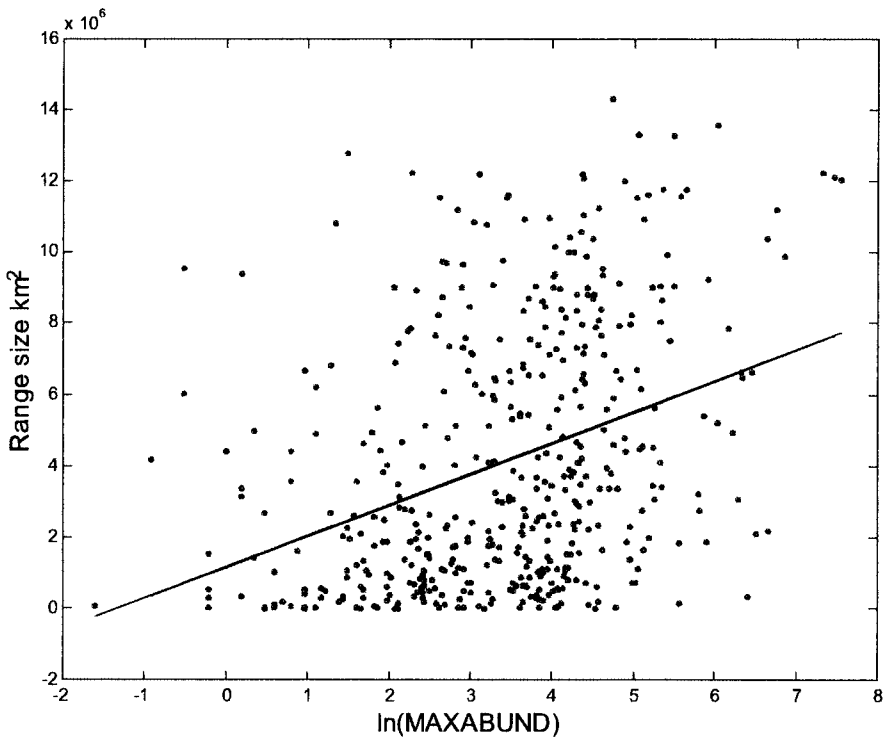

Fig. 6. Plot of species range size versus ln(abundance). The theory predicts an increasing linear relationship. The data from the Breeding Bird Survey clearly show a positive relationship (slope = 1.15×10^6, $P < 0.001$), but there is significant scatter ($r^2 = 0.13$).

1. *For independently derived, a priori parameters, the SPAR model produces a SPAR similar to that found in nature*. We estimated all parameters of the SPAR model using primarily BBS data (in particular, no parameters were chosen to maximize goodness-of-fit of the curve or even using the data we were fitting to (see Appendix 1 for details). With these *a priori* parameters, the SPAR model produces a SPAR that is nearly identical to Preston's (1960) classic empirical SPAR for North American birds (Fig. 7). Thus, the model correctly generates the correct slope (z), intercept (c) and location and degree of curvilinearity. The idea of calculating a species–area curve from randomly placed species ranges is not novel, but this is the first time that such a simulation has produced c and z values similar to those in nature without a free (fitting) parameter (cf. Leitner and Rosenzweig, 1997; Maurer, 1999).
2. *The shapes of the local and global abundance distributions are very similar*. Our analytical model for SADs made the novel prediction that the shape (degree of convexity) of the local and global SADs should be the same. They are – when fitted with a power distribution, the global abundances (here we used AVGABUND, since it is not confounded with the number of peaks) have a c parameter of 0.293, while the local abundances have an average c of 0.289 with 95% of the values falling between 0.204 and 0.380. We use the power distribution here because it has a single parameter, c, that directly corresponds to shape only (as opposed to scale, etc.). The fact that the local abundances across 1000 sites

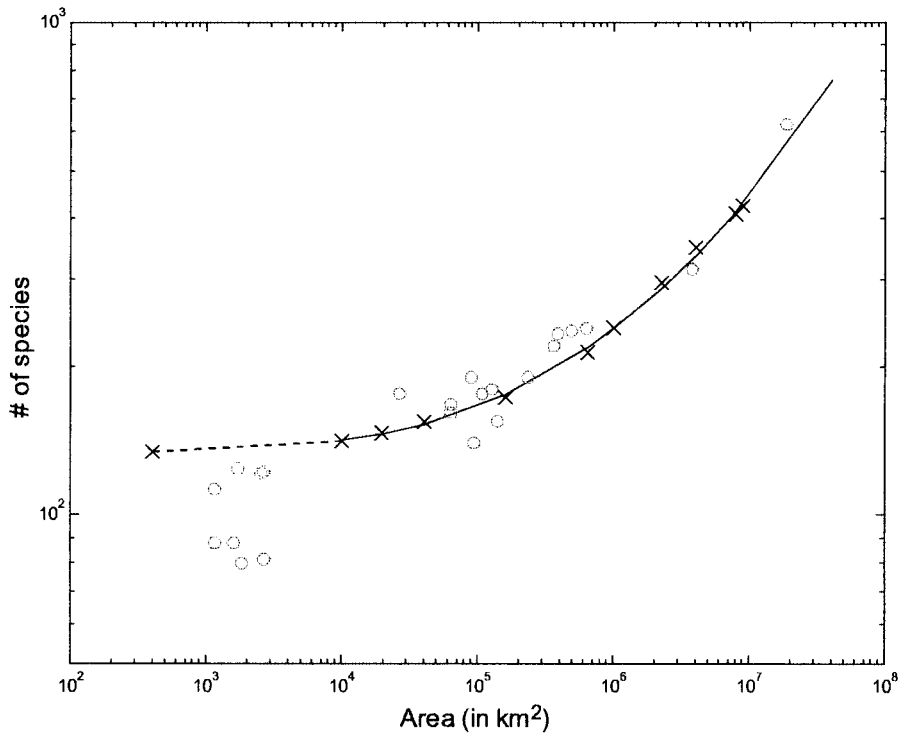

Fig. 7. Predicted versus empirical SPAR. This graph shows the simulated SPAR model (× symbols with line showing an average across 10 Monte Carlo simulations) versus Preston's (1960) data for birds of North America (○ symbols). The line is nearly horizontal between the leftmost two points as described in the text. Note that in this same region, the data for scales smaller than 10^4 km^2 fall below the line. This is because SPARs at small scales are driven by sampling effects and habitat heterogeneity, which are not included in our model. The rest of the line is a fit of an exponential curve in log–log space. Except at very small scales, we can see that the simulated data come extremely close to the observed data, especially given that no fitting parameters were used. The simulation correctly predicts the *c* value (intercept), *z* value (slope) and region and degree of curvilinearity.

all have a similar *c* value, which, in turn, is similar to the *c* value for global abundance, is striking and suggests that global abundances may in fact drive local abundances to some degree and through some unknown mechanism (e.g. source–sink population dynamics). This is contrary to current thinking, which assumes that global abundances are merely the sum of local abundances.

3. *Local abundance increases with proximity to an abundance peak and with global abundance*. We sampled 1000 routes and looked at the local abundances at each of these routes. This gave us a total of 76,581 species–route combinations. In particular, log(local abundance) was correlated with global abundance (log(MAXBUND)) with a mean Pearson correlation $r = 0.47$, and also with distance from peak (log(DIST)) with $r = -0.41$. Non-parametric Spearman rank correlations and other measures of global abundance and distance from peak abundance gave very similar results.

Over the 1000 routes, 95% of the Pearson r values are in the range 0.19–0.66 for log(MAXABUND) and 0.00 to –0.65 for %DIST. All but eight of the 1000 routes are significant at the $P < 0.001$ level for log(MAXABUND) and all but 26 routes are significant at the $P < 0.005$ level for log(DIST). Fine-scale spatial autocorrelation could be partly responsible, but the average distance for a given species in a given route from its nearest peak is 682 km with a standard deviation of 694 km. Thus, the spatial autocorrelation would have to be acting over very long distances. Indeed, spatial correlation would have to be acting over the scale of the entire species range, just as we have proposed in our model.

4. *Our theory explains a large proportion of the variance in local relative rank*. This is clearly related to the previous test. Rank of a species (where 1 = most abundant) can be predicted by:

$$\text{RANK} = c + c_1 \log(\text{MAXABUND}) + c_2 \log(\text{DIST}) + \text{noise} \quad (3)$$

The average R^2 (proportion of variance explained by the two variables) for 1000 routes was 87%, with 95% of the values falling in the range 76–93%. All 1000 routes were significant at the $P < 0.001$ level. Thus, local community importance of a species is strongly determined by (or at least correlated with) events occurring hundreds and thousands of kilometres away.

All the above results are conservative for at least two reasons. First, the algorithms for identifying peaks and range boundaries were relatively crude. We found that humans looking at range maps produced data that provided even stronger correlations than those reported, but it was not possible to do this in a scaleable or patently objective manner. Second, the BBS data are often at the wrong spatial scale for the process. In the mountainous west, the Gaussian distribution of abundance across an altitudinal gradient probably drives local abundance more than a Gaussian distribution of abundance across a species range. This issue is confirmed by the fact that 48 of the 50 routes with the lowest proportion of variance explained are in mountainous regions of the west. Shoreline/inland and urban/rural gradients might also affect these results.

DISCUSSION

It is important to recognize the limits of the model. It fails to explain species–area relationships (SPARs) on small scales where sampling and habitat heterogeneity are dominant factors. It also fails to explain between a quarter to a half of the variation in local rank and abundances in species abundance distributions (SADs). The remaining variation is presumably due to local processes, including abiotic and biotic factors, which we do not model.

Within these limits, the model has great explanatory power. This leads us to examine what mechanisms (proximate or ultimate) underlie our three premises. So, having spent time demonstrating how our three premises build up to produce four patterns, we now reverse direction and explore the mechanisms behind the premises. Despite the fact that the premises are well documented, their causes are still poorly known.

The first premise, independence of range locations, is not surprising due to a statistical argument – the law of rare events (Taylor and Karlin, 1998). This law suggests that a great

many things should be distributed across space according to a Poisson model, much like the central limit theorem suggests that normal distributions should be common.

We do not understand the cause of interspecific variation in global abundance (Premise 2). There is evidence that species usually retain their rare or abundant status over long periods of time (McGowan and Walker, 1985; Boucot, 1996; Brett *et al.*, 1996; Hadly and Maurer, 2001), albeit subject to major community restructuring events (such as the retreat of the glaciers). This suggests that relative global abundance might be an inherent property of the species. As shown here and elsewhere (Brown, 1995; Gaston and Blackburn, 2000), it is strongly correlated with range size. It also seems to be correlated with high rates of population growth (Kunin and Gaston, 1997; Duncan *et al.*, 1999). We can rule out one common explanation, body size. Global abundance is not strongly correlated with body size. Regression of body size versus global abundance on BBS data is significant at $P < 0.0001$ with a slope of –0.11, but the regression explains only 2.75% of the variance (see also Brown and Maurer, 1987). Those traits that have been found to correlate positively with large ranges all appear to fall into the general category of 'weediness' – high rates of and allocation to reproduction, fast lives (short time to maturity, early death), high dispersal abilities, living in marginal (often disturbed, low diversity) habitats, and being competitively subordinate (Glazier, 1980; Gaston and Kunin, 1997; Rosenzweig and Lomolino, 1997; Glazier and Eckert, 2002).

Neither do ecologists understand what causes the third premise: peak-and-tail spatial variation in abundance across a range (Brown *et al.*, 1996). One of the authors explores this pattern and its possible mechanisms in considerable detail in another paper (B.J. McGill, submitted). To date, three main theories have been proposed (Hengeveld and Haeck, 1981; Brown, 1984): one based on physiological response surfaces, one based on dispersal and one based on the Hutchinsonian niche and the central limit theorem. Physiological response surfaces plot some component of fitness (such as fecundity) against an abiotic factor such as temperature, light or nutrients. Gause (1932) presented some early physiological response surfaces and suggested that they were Gaussian (bell-curved) in shape, which explained the peak-and-tail pattern. Unfortunately, this has not held up. In the modern literature, most physiological response surfaces are either Monod/Michaelis-Menton (Botkin, 1993; Pacala and Kinzig, 2002) or parabolic (Botkin, 1993; Guttierez, 1996). These explain the peaks and sharp drop-offs, but not the tails. The second proposed mechanism is dispersal, which is modelled by diffusion equations. The solution to the diffusion equation gives a travelling wave, which is Gaussian (Skellam, 1951; Turchin, 1998). However, this assumes the ability to reach infinite abundance at one site and to grow in space without limits. Adding a carrying capacity and finite boundaries gives a solution with a plateau and a sharp drop-off but with no tails (Skellam, 1951; Kot, 2001). Recent work (B.J. McGill, submitted) has shown that the addition of a long-tailed dispersal kernel (Clark, 1998) can create tails that possess a source–sink structure at the scale of the species range (the tails occur in regions where fitness is below 1.0). The time-scale of this large-scale source–sink dynamic may be several generations (Maurer and Villard, 1994). Source–sink dynamics at the scale of the species range is an old idea (Grinnell, 1904; Maurer and Villard, 1994; Lawton, 1996; Maurer, 1999; Pulliam, 2000). The third theory points out that the Hutchinsonian niche has many dimensions, and hence the physiological response surfaces along each dimension are added or multiplied together, allowing an argument based on the central limit theorem to cause the tails (Brown, 1984; Brown *et al.*, 1995). Recent work (B.J. McGill, submitted) suggests a fourth mechanism, trade-offs in several components of fitness along a single

environmental gradient. For example, the trade-off between competitive dominance (presumably leading to increased food intake and hence fecundity) and environmental tolerance (leading to increased survival) can combine to create a peak-and-tail structure. Testing these mechanisms is highly challenging because of the spatial scales involved. Recent work offers hope though. Laboratory physiology is beginning to be applied to problems at the scale of species ranges (Davis *et al.*, 1998), and improved molecular techniques provide hope for tracking large-scale movements of individuals (Clobert *et al.*, 2001). Ultimately, theory and testing about these mechanisms must be traced to population dynamics (Maurer, 1994; Maurer and Taper, 2002). In all likelihood, all four of these mechanisms are involved in creating the peak-and-tail pattern.

There are several implications of this paper for conservation biology. If we are correct that local SAD curves are structured primarily by where the community is located within the independent ranges of the species, and not primarily by interspecific interactions or local abiotic factors, then the design of reserves will have to be approached very differently. Current analysis of the siting of reserves pays little if any attention to the position of a reserve within the species range. The consequences of ignoring this depend on the mechanisms that structure abundance across a species range (Premise 3). If the peak-and-tail pattern is created purely by physiological responses, then limiting a species to a small fraction of its former range might have no effects. But if dispersal plays a significant role in structuring abundances across a range, then this would have radical effects. In particular, large portions of the range would be sinks, and the species would be doomed in the long run if a reserve was located within sinks and the sources were eliminated (Grinnell, 1904; Lawton, 1993; Maurer and Villard, 1994; Lawton, 1996). The high variation in abundance across a species range also suggests that we need to pay more attention to the abundance of species and not just the presence or absence of species when we design reserves (Brown *et al.*, 1995). Much current work on locating reserves is based solely on presence/absence data (Scott *et al.*, 1993; Jennings *et al.*, 1997). Finally, the abrupt transition in the scale at which our theory fails to explain and then explains the SPAR may be of use (Fig. 7). Specifically, this scale (about 10^4 km^2) represents for North American birds the scale at which a reserve is large enough to contain all the species whose species ranges intersect the reserve. As discussed, smaller scales do not include either enough individuals or enough habitat heterogeneity to provide a comprehensive sample of all the species that are supported in the region. Determination of this scale for various taxa could provide a very useful guideline for desirable reserve sizes. Admittedly, the resulting reserve size, at least for birds, is quite large (Yellowstone National Park is about this size at approximately 9000 km^2), but it may not be a coincidence that these very large reserves play such a vital role in conservation.

CONCLUSIONS

In summary, we have presented a model using three well-documented premises, which produces four qualitative, well-known patterns as well as three novel, quantitative predictions. We have endeavoured to provide a strong test of this theory. The main implication of the success of our model is the need to understand what processes underlie our three premises better. We hope that this work stimulates more research directed towards understanding the processes driving variation in abundances across ranges and in global abundance of species.

ACKNOWLEDGEMENTS

We first thank all the volunteers who spent many hours collecting the BBS data and the professionals at Patuxent Wildlife Research Center who have put the data into an accessible, high-quality format. We also thank Mike Rosenzweig and the students in his lab for stimulating our thinking in this area. We thank people who kindly read earlier versions of this manuscript: Will Turner, Yaron Ziv, Henrique Pereira, Sean Connolly, Mike Rosenzweig and Brian Enquist. B.J.M. thanks the NSF and the Flinn Foundation Biomath Program at the University of Arizona for funding. C.D.C. thanks the NSF RTG for Biodiversity at the University of Arizona.

REFERENCES

Austin, M.P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio*, **69**: 35–45.

Botkin, D.B. 1993. *Forest Dynamics: An Ecological Model*. Oxford: Oxford University Press.

Boucot, A.J. 1996. Epilogue. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **127**: 339–359.

Brett, C.E., Ivany, L.C. and Schopf, K.M. 1996. Coordinated stasis: an overview. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **127**: 1–20.

Brown, J.H. 1984. On the relationship between abundance and distribution of species. *Am. Nat.*, **124**: 255–279.

Brown, J.H. 1995. *Macroecology*. Chicago, IL: University of Chicago Press.

Brown, J.H. and Maurer, B.A. 1987. Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *Am. Nat.*, **130**: 1–17.

Brown, J.H., Mehlman, D.H. and Stevens, G.C. 1995. Spatial variation in abundance. *Ecology*, **76**: 2028–2043.

Brown, J.H., Stevens, G.C. and Kaufman, D.M. 1996. The geographic range: size, shape boundaries, and internal structure. *Annu. Rev. Ecol. Syst.*, **27**: 597–623.

Clark, J.S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *Am. Nat.*, **152**: 204–224.

Clobert, J., Danchin, E., Dhondt, A.A. and Nichols, J.D. 2001. *Dispersal*. Oxford: Oxford University Press.

Condit, R., Pitman, N. and Leigh, E.G. *et al.* 2002. Beta-diversity in tropical forest trees. *Science*, **295**: 666–669.

Darwin, C. 1859. *On the Origin of Species*. London. Clows & Sons.

Davis, A.J., Lawton, J.H., Shorrocks, B. and Jenkinson, L.S. 1998. Individualistic species responses invalidate simple physiological models of community dynamics under global environmental change. *J. Anim. Ecol.*, **67**: 600–612.

Duncan, R.P., Blackburn, T.M. and Veltman, C.J. 1999. Determinants of geographical range sizes: a test using introduced New Zealand birds. *J. Anim. Ecol.*, **68**: 963–975.

Enquist, B.J., Jordan, M.A. and Brown, J.H. 1995. Connections between ecology, biogeography, and paleobiology: relationship between local abundance and geographic distribution in fossil and recent molluscs. *Evol. Ecol.*, **9**: 586–604.

Evans, M., Hastings, N. and Peacock, B. 1993. *Statistical Distributions*. New York: Wiley.

Gaston, K.J. 1994. *Rarity: Population and Community Biology Series*, Vol. 13. London: Chapman & Hall.

Gaston, K.J. 1996a. Global scale macroecology: interactions between population size, geographic range size and body size in the Anseriformes. *J. Anim. Ecol.*, **65**: 701–714.

Gaston, K.J. 1996b. The multiple forms of the interspecific abundance–distribution relationship. *Oikos*, **76**: 211–220.

Gaston, K.J. and Blackburn, T.M. 2000. *Pattern and Process in Macroecology*. Oxford: Blackwell Science.

Gaston, K.J. and Kunin, W.E. 1997. Rare–common differences: an overview. In *The Biology of Rarity: Causes and Consequences of Rare–Common Differences* (W.E. Kunin and K.J. Gaston, eds), pp. 12–29. London: Chapman & Hall.

Gauch, H.G.J. and Whittaker, R.H. 1972. Coencline simulation. *Ecology*, **53**: 446–451.

Gause, G.F. 1932. Ecology of populations. *Quart. Rev. Biol.*, **VII**: 27–46.

Glazier, D.S. 1980. Ecological shifts and the evolution of geographically restricted species of North American *Permyscus* (mice). *J. Biogeogr.*, **7**: 63–83.

Glazier, D.S. and Eckert, S.E. 2002. Competitive ability, body size and geographical range size in small mammals. *J. Biogeogr.*, **29**: 81–92.

Grinnell, J. 1904. The origin and distribution of the Chestnut-backed Chickadee. *Auk*, **21**: 364–378.

Grinnell, J. 1922. The role of the 'accidental'. *Auk*, **39**: 373–380.

Guttierez, A.P. 1996. *Applied Population Ecology: A Supply–Demand Approach*. New York: Wiley.

Hadly, E.A. and Maurer, B.A. 2001. Spatial and temporal patterns of species diversity in montane mammal communities of western North America. *Evol. Ecol. Res.*, **3**: 477–486.

Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, **38**: 210–221.

Hanski, I. and Gyllenberg, M. 1997. Uniting two general patterns in the distribution of species. *Science*, **275**: 397–400.

Harte, J., Kinzig, A.P. and Green, J. 1999. Self-similarity in the distribution and abundance of species. *Science*, **284**: 334–336.

Hengeveld, R. and Haeck, J. 1981. The distribution of abundance. II. Models and implications. *Proc. Koninklijke Nederlandse Akadamie Van Wetenschappen C*, **84**: 257–284.

Hoagland, B.W. and Collins, S.L. 1997. Gradient models, gradient analysis, and hierarchical structure in plant communities. *Oikos*, **78**: 23–30.

Hubbell, S.P. 2001. *A Unified Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.

Jennings, M.D., Csuti, B. and Scott, J.M. 1997. Wildlife habitat relationship models: distribution and abundance. *Conserv. Biol.*, **11**: 1271–1272.

Kot, M. 2001. *Elements of Mathematical Biology*. Cambridge: Cambridge University Press.

Kunin, W.E. and Gaston, K.J. 1997. *The Biology of Rarity: Causes and Consequences of Rare–Common Differences*. London: Chapman & Hall.

Lawton, J.H. 1993. Range, population abundance and conservation. *Trends Ecol. Evol.*, **8**: 409–413.

Lawton, J.H. 1996. Population abundances, geographic ranges and conservation: 1994 Witherby Lecture. *Bird Study*, **43**: 3–19.

Leitner, W.A. and Rosenzweig, M.L. 1997. Nested species–area curves and stochastic sampling: a new theory. *Oikos*, **79**: 503–512.

Lindgren, B.W. 1976. *Statistical Theory*. New York: Macmillan.

MacArthur, R.H. 1972. *Geographical Ecology: Patterns in the Distribution of Species*. Princeton, NJ: Princeton University Press.

Maurer, B.A. 1994. *Geographical Population Analysis: Tools for the Analysis of Biodiversity*. Boston, MA: Blackwell Science.

Maurer, B.A. 1999. *Untangling Ecological Complexity*. Chicago, IL: University of Chicago Press.

Maurer, B.A. and Taper, M.L. 2002. Connecting geographical distributions with population processes. *Ecol. Lett.*, **5**: 223–231.

Maurer, B.A. and Villard, M.A. 1994. Geographic variation in abundance of North American birds. *Research and Exploration*, **10**: 306–317.

May, R.M. 1975. Patterns of species abundance and diversity. In *Ecology and Evolution of Communities* (M.L. Cody and J.M. Diamond, eds), pp. 81–120. Cambridge, MA: Belknap Press of Harvard University Press.

McGowan, J.A. and Walker, P.W. 1985. Dominance and diversity maintenance in an oceanic ecosystem. *Ecol. Monogr.*, **55**: 103–118.

Minchin, P.R. 1989. Montane vegetation of Mt Field massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio*, **83**: 97–110.

Murray, B.R., Rice, B.L. and Keith, D.A. *et al.* 1999. Species in the tail of rank-abundance curves. *Ecology*, **80**: 1806–1816.

Pacala, S.W. and Kinzig, A.P. 2002. Introduction to theory and the common ecosystem model. In *The Functional Consequences of Biodiversity* (D. Tilman and P. Kareiva, eds), pp. 169–174. Princeton, NJ: Princeton University Press.

Patuxent Wildlife Research Center. 2001. *Breeding Bird Survey FTP Site, Patuxent Wildlife Research Center*. Laurel, MD: PWRC.

Pielou, E.C. 1977. *Mathematical Ecology*. New York: Wiley.

Preston, F.W. 1948. The commonness and rarity of species. *Ecology*, **29**: 254–283.

Preston, F.W. 1960. Time and space and the variation of species. *Ecology*, **41**: 611–627.

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

Price, G.R. 1970. Selection and covariance. *Nature*, **227**: 520–521.

Price, J., Droege, S. and Price, A. 1995. *The Summer Atlas of North American Birds*. San Diego, CA: Academic Press.

Pulliam, H.R. 2000. On the relationship between niche and distribution. *Ecol. Lett.*, **3**: 349–361.

Raunkiaer, C. 1934. *The Life Forms of Plants and Statistical Plant Geography; Begin the Collected Papers of C. Raunkiaer*. Oxford: Clarendon Press.

Robbins, C.S., Bystrak, D. and Geissler, P.H. 1986. *The Breeding Bird Survey: Its First Fifteen Years, 1965–1979.* Resource Publication 157. Washington, DC: US Department of the Interior, Fish and Wildlife Service.

Rosenzweig, M.L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

Rosenzweig, M.L. and Lomolino, M.V. 1997. Who gets the short bits of the broken stick. In *The Biology of Rarity: Causes and Consequences of Rare–Common Differences* (W.E. Kunin and K.J. Gaston, eds), pp. 63–90. London: Chapman & Hall.

Sagarin, R.D. and Gaines, S.D. 2002. The 'abundant' centre distribution: to what extent is it a biogeographical rule? *Ecol. Lett.*, **5**: 137–147.

Sauer, J.R., Hines, J.E., Gough, G., Thomas, I. and Peterjohn, B.G. 1997. *The North American Breeding Bird Survey Results and Analysis, Patuxent Wildlife Research Center*. Laurel, MD: PWRC.

Scott, J.M., Davis, F. and Csuti, B. *et al*. 1993. Gap analysis – a geographic approach to protection of biological diversity. *Wildl. Monogr.*, **123**: 1–41.

Shipley, B. and Keddy, P.A. 1987. The individualistic community-unit concepts as falsifiable hypotheses. *Vegetatio*, **69**: 47–55.

Sibley, D.A. 2000. *National Audobon Society: The Sibley Guide to Birds*. New York: Alfred A. Knopf.

Skellam, J.G. 1951. Random dispersal in theoretical populations. *Biometrika*, **38**: 196–218.

Stoyan, D. and Stoyan, H. 1994. *Fractals, Random Shapes and Point Fields*. Chichester: Wiley.

Taylor, H.M. and Karlin, S. 1998. *An Introduction to Stochastic Modelling*. San Diego, CA: Academic Press.

Turchin, P. 1998. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sunderland, MA: Sinauer Associates.

Whittaker, R.H. 1951. A criticism of the plant association and climatic climax concepts. *Northwest Sci.*, **25**: 17–31.

Whittaker, R.H. 1965. Dominance and diversity in land plant communities. *Science*, **147**: 250–260.

Wilkinson, L. 1997. *SYSTAT 7.0 Statistics*. Chicago, IL: SPSS Inc.

APPENDIX 1: METHODS FOR SPAR

To show analytically that our model produces SPARs would require new results in stochastic geometry, something we have been unable to accomplish.* Thus, we rely entirely on implementing a Monte Carlo simulation of the three mathematical premises and the one technical assumption described in the paper. In this description of the simulations, capital letters denote a model variable, with bold face indicating an input variable. See Fig. A1 for a visual description of the model. We model the spatial extent as a square of size **WIDTH** × **WIDTH** (the edges are not wrapped into a torus as in some spatial simulations). Let **S** be the number of species occurring in this area. We place **S** different points, μ_i, randomly in this square according to a two-dimensional Poisson process and let these points represent the centre of the range for each species.

We then need to determine the RADIUS_i of the range centred at each point. We can do this in one of two ways. The simplest is to randomly resample from the actual ranges sizes observed in the

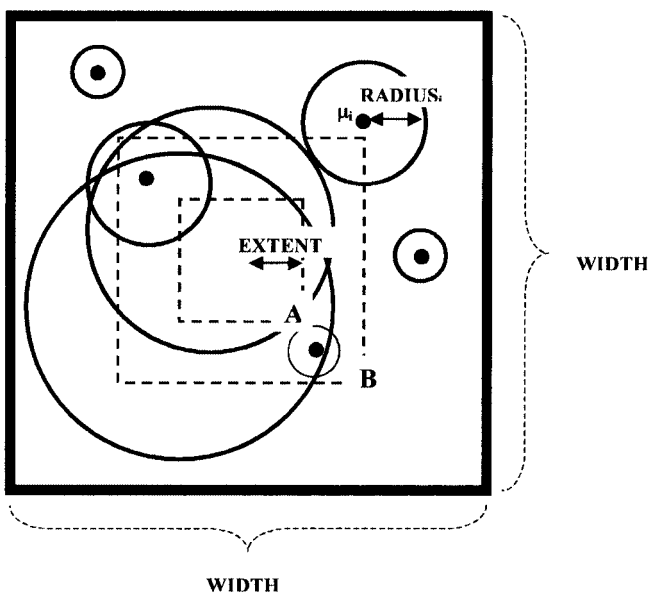

Fig. A1. A pictorial representation of the SPAR simulation model and its parameters. A continent of dimensions WIDTH × WIDTH is created. Within this continent, the centres, μ_i, of species are placed down randomly using a Poisson process. For each centre, a RADIUS is randomly sampled. One alternative for the distributions of the radii is to resample from an empirical distribution of range sizes. Another alternative is to use equation (2) and specify distributions for NMAX_i, σ_i and the value of NMIN. See the text for descriptions of the distribution of radii. Ranges are treated as circular. A progressively larger series of boxes is drawn, each with a size of EXTENT × EXTENT where the maximum EXTENT = WIDTH/2. In the above figure, box A has a species diversity of 3 (inside of two ranges and intersects a third). Box B has a species diversity of 5 (the previous 3, one whose range is completely included and one whose range it intersects). The total diversity of the continent, *S*, is 7. *S* is an input parameter to the model.

* While this paper was in final revision, a manuscript was shared with one of us by Andrew Allen and Ethan White at the University of New Mexico providing an analytical solution to the model described in this section, which we had to solve by computer modelling. Their paper also appears in this issue (pp. 493–499).

BBS – treating them as circles and calculating RADIUS = $\sqrt{}$(rangearea/π). Alternatively, we can use equation (2). Then we must specify the distributions of σ_i and of NMAX_i and the value of NMIN. We use a power distribution to represent the distribution of σ_i, since range distributions are right-skewed on a log scale. This gives us two parameters, $\mathbf{C}_\sigma$ (the exponent or shape of the power distribution) and $\sigma_{\mathbf{max}}$ (the largest value of σ). We model NMAX_i as varying according to either a power distribution or a log-normal distribution, as in the SAD model.

We performed sensitivity analysis on all of these parameters. In all cases, power-law SPARs were produced. The z values varied over a relatively biologically reasonable range of 0.05–0.25. Only one parameter has a large effect on the z values: the spatial scale of the RADIUS's (although not the exact distributional shape). Resampling ensures that we are using exactly what is found in nature. In the non-resampling model, the parameter $\sigma_{\mathbf{max}}$ and those that drive the distribution of NMAX are the most critical. The intercept of the curve was sensitive primarily to the ratio of the number of species **S** to the area (**WIDTH** $\times$ **WIDTH**).

To demonstrate that our model is robust to deviations in premise, we also ran Monte Carlo simulations with all of the alternatives described in Table 2. In all cases, results were very similar with these modified assumptions, indicating that the SPAR model is robust to variation in the premises.

We derived the independent (i.e. non-curve-fitting) estimates of these parameters from the BBS data as follows. We took **WIDTH** to be 3000 km – giving an area of 9 million km^2, close to the size of the continental USA and the part of southern Canada captured in the BBS. Surprisingly, there are a great variety of estimates of the number of species of birds in this area, **S**, depending on time of year and intensity of sampling. Since we were going to compare with Preston's data set (Preston, 1960), we needed to find an estimate of bird diversity in our target region which uses the same standards as Preston: the number of birds which regularly breed there (Preston defines this as 9 out of 10 years as estimated by an expert). We obtained an estimate of this by looking at Sibley (2000). Sibley covers all birds encountered, not just those reliably breeding, so an adjustment is required. Preston lists 625 species living in the Nearctic (Canada and the USA, including Alaska but not Hawaii). Looking at Sibley's range maps, we count 549 species which summer or live year-round in the BBS area (continental USA and southern Canada). Sibley includes 810 species living in a region which matches Preston's Nearctic. Thus we may assume that 549/810 = 67.8% of all species encountered in the Nearctic are encountered in the target BBS area. Applying this to Preston's 625 species that regularly breed in the Nearctic, we get 67.8% $\times$ 625 = 424 species that regularly breed in the target area. As a sanity check, we also perform an interpolation on Preston's curve, by taking the three points closest to 9×10^6 km^2 on his nearctic SPAR and fit a quadratic polynomial to estimate the species abundance in North America arriving at 466.7 species. This result is fairly close and sensitivity analysis suggests that the resulting SPAR is not too different for either of these estimates. We use the estimate **S** = 424 since it was derived independently of that portion of Preston's data which we are comparing our predicted curve to.

We derived independent estimates for the distribution of RADIUS as follows. For the resampling method of determining RADIUS, no parameters were needed. We simply used the range sizes (defined as the convex hulls around all routes where the species was found converted into area in km^2). For the method using equation (2), we used NMAX_i distributed as a log-normal with a mean of 3 and standard deviation of 1.3 (MLE estimates for the global distribution of average abundance in the BBS). The parameter $\sigma_{\mathbf{max}}$ represents the distance from the peak to the inflection point of rapid drop off in abundance on the largest range. We used $\sigma_{\mathbf{max}}$ = 1044, derived by estimating the area where the Red-winged blackbird (species with the largest fully occupied range) had an abundance > 50 (abundance where rapid drop off in abundance seemed to occur) by multiplying the number of routes where the abundance was > 50 times the average area/good route ($\approx$ 7000 km^2) in the BBS (see Appendix 3 for definition of a good route). We used a power distribution with an exponent $\mathbf{C}_\sigma$ = 0.5 (MLE estimate for BBS range sizes) for the variation in σ. We used **NMIN** = 0.25, the middle of three values used by Price *et al*. (1995).

APPENDIX 2: SAD ANALYTICAL MODEL

The key trick in calculating the distribution of abundance, N_i, over all species, i, at a given location, x, in the interval [0,1] is to note that N_i is symmetric in x and μ_i (see equation 1). Thus sampling all species at one point is identical to sampling one species at all points (where the sampling is done with NMAX_i and μ_i retained as random variables).

This can now be solved using simple calculus and probability definitions (Lindgren, 1976, pp. 85–87). Let F be a cumulative distribution function. Then the distribution of interspecific abundances, N_i, at a point (say zero, without loss of generality) = the distribution of abundances of one species where NMAX and μ are now random variables:

$$F_{\mathrm{N}}(n) = \Pr\{N \le n\} = \Pr\{\mathrm{NMAX}\exp(-\mu^2/2\sigma^2) \le n\} = \Pr\{\mathrm{NMAX} \le n\exp(\mu^2/2\sigma^2)\}$$

if we let R represent the region where $0 \le \mu \le 1$ and $0 \le \mathrm{NMAX} \le n\exp(\mu^2/\sigma^2)$, then

$$F_{\mathrm{N}}(n) = \iint_{\mathrm{R}} f_{\mathrm{NMAX},\mu}(\mathrm{NMAX}, \mu)\, d\mathrm{NMAX}\, d\mu = \iint_{\mathrm{R}} f_{\mathrm{NMAX}}(x) f_{\mu}(y)\, d\mathrm{NMAX}\, d\mu$$

(due to independence of NMAX and μ), so

$$F_{\mathrm{N}}(n) = \int_0^1 \int_0^{n\exp(\mu^2/2\sigma^2)} c\mathrm{NMAX}^{c-1} \bullet 1\, d\mathrm{NMAX}\, d\mu = \int_0^1 \mathrm{NMAX}^c \Big|_0^{n\exp(\mu^2/2\sigma^2)} d\mu = n^c \int_0^1 \exp(c\mu^2/2\sigma^2)\, d\mu$$

or in conclusion

$$F_{\mathrm{N}}(n) = C(c,\sigma)\, n^c$$

where $C(c,\sigma)$ is a normalization constant depending in an increasing fashion on c and in a decreasing fashion on σ.

Hence, N is distributed as a power distribution with shape parameter c. Note that if we start out assuming that NMAX has a distribution other than a power distribution, we do not get the power distribution for local abundance. A similar analytical approach with a uniform distribution of global abundances also gives a hollow curve, as does Monte Carlo simulations with a log-normal distribution.

APPENDIX 3: METHODS FOR BBS

The Breeding Bird Survey (Robbins *et al.*, 1986; Sauer *et al.*, 1997; Patuxent Wildlife Research Center, 2001) is conducted annually during the peak bird breeding season (May and June) and consists of 50 separate three-minute point counts over a 24.5 mile long route. Volunteers collect the data annually at thousands of different sites across the entire continental USA and southern Canada. It is unique in its spatial extent, which is important to this paper. The methods used do not give absolute species abundances, only relative, but this is adequate for the purposes of this paper, since all uses of the data are comparisons to other sites or species.

The 1966–2000 data set was downloaded on the internet via ftp (Patuxent Wildlife Research Center, 2001). The years 1996–2000 were averaged across time to remove sampling error and other noise while still using a small enough temporal period to avoid problems with long-term trends of changing abundance. This gave us a data set with abundance at 4852 routes for 1186 species. The BBS assigns a route quality to each route/year pair. We used only routes that the BBS administrators considered good for all 5 years, reducing us to 1401 routes. Thirty-seven species in the data are aggregate species

(due to taxonomic splitting after data were collected) and a number of others were not observed in our 5-year period. We eliminated these species, leaving us 564 species. Three measures of global abundance were calculated: TOTABUND = sum of abundances at all good routes for a species, MAXABUND = the largest abundance over any good route and AVGABUND = the average abundance over all good routes where the species is found (abundance > 0). The mathematical model is based on MAXABUND, the height of the peak, and we use this in regressions and correlations. However, since MAXABUND is not independent of the number of peaks, we use AVGABUND to represent the shape of the distribution of global abundances. We also calculated two measures of range size: NUMRTS was the number of good routes at which the species was found, RANGEAREA was the square kilometres inside the convex hull surrounding all good routes where the species was found.

Most of our results come from randomly selecting 1000 good routes and, for each route, calculating:

- The observed cumulative distribution function (CDF) of bird abundances.
- The goodness-of-fit of the observed CDF with the CDF of several different theoretical distributions using two methods of calculating non-linear R^2 (proportion of variance explained and square of correlation coefficient of histograms) and the Kolmogorov-Smirnov statistic.
- For each species we calculated the nearest peak, the distance in kilometres, DIST, to the nearest peak, and %DIST (the distance to the nearest peak divided by the length of a line passing from the peak through the particular route to the edge of the range as calculated by the convex hull method). Peaks were defined heuristically as any route with an abundance greater than 80% of the highest observed abundance over all routes (MAXABUND) on a log scale (exp(0.80 × log(MAXABUND)).

All variables were log-transformed except %DIST. Both Pearson and Spearman correlations were calculated and found to be similar. We present the Pearson correlations as they have a stronger interpretation. Results did not vary substantially regardless of which measures of global abundance or distance were used. In most cases, the statistics presented represent the mean and the 2.5th and 97.5th percentiles (i.e. includes 95% of routes) over the 1000 routes examined.