

# Review Paper The interpretation of biological surveys

# **Graham Bell**

Redpath Museum and Biology Department, McGill University, Montreal, Québec, H3A 1B1 Canada (graham.bell@mcgill.ca)

Biological surveys provide the raw material for assembling ecological patterns. These include the properties of parameters such as range, abundance, dispersion, evenness and diversity; the relationships between these parameters; the relationship between geographical distributions and landscape structure; and the co-occurrence of species. These patterns have often been used in the past to evaluate the role of ecological processes in structuring natural communities. In this paper, I investigate the patterns produced by simple neutral community models (NCMs) and compare them with the output of systematic biological surveys. The NCM generates qualitatively, and in some cases quantitatively, the same patterns as the survey data. It therefore provides a satisfactory general theory of diversity and distribution, although what patterns can be used to distinguish neutral from adaptationist interpretations of communities, or even whether such patterns exist, remains unclear.

Keywords: species richness; evenness; abundance; range; diversity; distribution

# 1. INTRODUCTION

Naturalists have compiled lists of species since the time of Aristotle; in the last century, systematic biological surveys have provided a rich source of data for community ecology. A single large-scale survey-of 1000 species, say, at 1000 sites-yields about as much information as the complete sequence of a small genome. Does it yield as much insight? The statistical analysis of biological surveys has often uncovered strong and consistent patterns that have given rise to a very large literature over the past 50 years (reviewed in Cody & Diamond 1975; Diamond & Case 1986; Gee & Giller 1987; Rosenzweig 1995; Weiher & Keddy 1999). These patterns have aroused much interest because they seem to offer the means of evaluating ecological mechanisms in circumstances where experimentation is impracticable. They have also aroused much controversy. In the past 2 years, it has been shown that many of the most general and best documented patterns are readily generated by neutral community models (NCMs) in which all individuals have the same demographic properties, regardless of the species they belong to (Bell 2001; Hubbell 2001). This might be literally true (the strong version of the theory) or it might furnish an appropriate null hypothesis despite the existence of variation among species (the weak version). In either case, one implication of this result is that the scrutiny of ecological patterns might not be a reliable technique for understanding ecological processes. This has divided opinion among ecologists (see Enquist et al. 2002; Whitfield 2002). Some regard neutral models as the most powerful general theory yet to emerge in community ecology, whereas others view them as being self-evidently wrong and therefore sterile. The purpose of this paper is to review the kinds of pattern that emerge from biological surveys and to evaluate the ability of the NCM to interpret them.

# (a) Neutral community models

The NCM represents a community in which all individuals have the same demographic properties: the same probabilities of birth, death and dispersal. It refers to communities of ecologically similar species where individuals compete for similar resources and in which all non-zero interactions are negative; it cannot be applied to trophically complex communities containing predators, prey, parasites, parasitoids, hosts or mutualists in which some interactions are positive. The main tenet of the NCM is that the bulk of the diversity found in a sample is neutral, that is, consists of species whose members have very similar demographic properties. It is not necessary that these properties be strictly identical (although they are in the models that I shall describe), but rather that any selection acting among species is weak relative to other processes such as dispersal.

In the NCM, it is individuals that have identical demographic properties. Species are not, in general, equivalent, because abundant and widespread species will usually differ systematically from rare or localized species. The theory of island biogeography (Macarthur & Wilson 1967) is an example of a model in which species have identical properties; in metapopulation theory (Levins 1969), populations have identical properties. Although these are neutral models of a sort, they are much less generally applicable than the NCM, where the properties of populations and species emerge from underlying demographic processes. The NCM is closely related to the neutral theory of molecular evolution (Kimura 1983), from which it differs chiefly in requiring that the total number of individuals at any given site has a fixed upper limit.

NCMs are the simplest representations of community dynamics. In general, they require specifying five parameters: the rates or probabilities of birth, death and dispersal; a rate of immigration or speciation; and a carrying capacity. Any model that involves local adaptation will require at least one more parameter (for example, a selection coefficient) and may require many more. NCMs are not random models, however. Restricted dispersal will readily lead to highly non-random patterns of distribution, association, abundance and diversity. The appropriate null hypothesis for testing theories that attempt to explain these phenomena is therefore likely to be a neutral model rather than a randomization test.

Two versions of the NCM have been developed. In the first, new species arise spontaneously, yielding a completely self-contained model at the expense of positing unexpectedly high speciation rates in many circumstances (Hubbell 2001). The second version allows new species to enter the community by immigration from an external pool (Bell 2000). This avoids the need to specify a mechanism for speciation, at the cost of restricting the model to ecological time-scales. This is the version used throughout this paper. If the rate of speciation or immigration is very low (of the order of the reciprocal of total number of individuals in the community or less) then the two versions give very similar output.

#### (b) The Mont St-Hilaire plant survey

I shall illustrate the application of NCMs to biological survey data with a systematic survey of sedges (Carex) in a large fragment of old-growth forest at Mont St-Hilaire, southern Quebec, Canada, 45.5° N, 73.1° W (http://www.mcgill.ca/gault). This forest has never been felled, cultivated or settled, and remains in a state that largely reflects the natural patterns of disturbance typical of this region since deglaciation. The presence or absence of 42 species of Carex was scored over an area of slightly more than 1000 ha (1 hectare =  $10^4 \text{ m}^2$ ) at a grain of 0.25 ha, yielding a total of 174 048 records. For analysing patterns of abundance and range, the records for Carex were agglomerated into 16 ha blocks, giving for each species a score between 0 and 64 that estimates 'cover', a common surrogate for abundance in plants that lack distinct individuality. A single integrated survey of a group of ecologically similar and phylogenetically related species thus provides a common platform for evaluating the neutral interpretation of survey data.

# 2. PATTERNS OF RANGE, ABUNDANCE AND DIVERSITY

#### (a) The species $\times$ sites matrix

The results of any survey can be expressed as a matrix of N species and M sites. The quantities that can be calculated directly from the species × sites matrix, and the symbols given to them, are shown in table 1. Each cell of this matrix records the occurrence (incidence) of a given species at a given site, either as binary data (Xij = 0 or 1, indicating the presence or absence of the *i*th species at the *i*th site) or as quantitative data (xij = number of individualsof the *i*th species at the *j*th site). There are therefore two sets of parameters that can be estimated from a survey, one referring to binary data and the other to quantitative data. Binary data are particularly important because the results of many surveys are given only as presence or absence, for example by shading squares on a grid. The row and column totals and variances then identify the fundamental parameters of abundance and diversity. I have designated each parameter as an upper-case letter for binary data and as the corresponding lower-case letter for quantitative data. Simberloff & Connor (1979) developed a similar matrix approach ('Q-mode and R-mode' analysis) for evaluating randomization tests of community models.

#### (b) Species richness and evenness

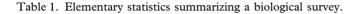
The column (site) total of binary data is S, the number of species recorded at a site, the simplest and most widely used measure of diversity, usually called species richness. For quantitative data, this total is density, the total number of records s from a site. The variances of these totals over the whole set of M sites,  $P_M$  and  $p_M$ , can be used to evaluate the significance and magnitude of variation in diversity and abundance. The column variance expresses taxonomic heterogeneity. For binary data this can be expressed as the binomial variance of species presence over the whole set of N species, Q = N(S/N)[(N - S)/N], although the distribution will not be binomial if species are correlated among sites. For quantitative data it is the parametric variance of abundance, q, which has often been called species evenness. Species richness and evenness have been reviewed by Kempton (1979); Washington (1984); Frontier (1985); Harper & Hawksworth (1994) and associated papers, and Smith & Wilson (1996), among others. The sampling properties of richness, evenness and related measures of diversity are outlined in a useful paper by Lande (1996), who also establishes the conditions for the additivity of within-site and among-site estimates of diversity.

#### (c) Range, abundance and dispersion

The row (species) total of binary data represents site occupancy R, one of the measures of range (see Gaston 1994b). The total for quantitative data is overall abundance r, the total number of individuals (or records) of a given species in the sample. The variances of these totals express the amount of variation of site occupancy  $Q_N$  or of abundance  $q_N$  among the whole set of N species. The overall evenness of the sample is the variance of row totals,  $q_{N_2}$  which will be minimal when all species are nearly equally frequent. It is inversely related to Simpson's Index, which is usually defined as the sum of squared frequencies, or, more generally, to a genetic variance as this would be defined (in a similar manner) by a population geneticist. The row variance is a measure of dispersion. For binary data this is the binomial variance of site occupancy among the whole set of M sites, P = M(R/M)[(M - R)/M]. For quantitative data it is the parametric variance of abundance over sites, p. Both dispersion p and evenness q will tend to increase with the number of records, so they are often more appropriately calculated as the corresponding variances of frequencies,  $\tilde{p}$  and  $\tilde{q}$ .

### (d) Overall diversity

Overall species richness within the survey area might be regulated by local processes of competition among the members of a nearly saturated community, or by regional processes of immigration from an external species pool (Ricklefs 1987). Attempts have been made to distinguish between these possibilities by showing that local species diversity either continues to rise in proportion to regional diversity, or approaches some limiting value, suggesting that ecological differentiation sets a limit to local diversity (Cornell & Lawton 1992). Most studies of this kind (reviewed by Srivastava (1999)) have found more or less linear plots, suggesting that dispersal is more important than competition in determining local diversity. This interpretation is hindered by the possibility that ecological



	species species n	m to <sup>at</sup> to <sup>at</sup> va <sup>tab</sup>	
type of data	binary	quantitative	characteristic of
row total	range (site occupancy) <i>R</i> : number of sites in which species occurs	abundance <i>r</i> : number of individuals of species over all sites	species
variance of row totals	variance of site occupancy among species $Q_N$	variance of abundance among species: overall evenness $q_N$	set of $N$ species
row variance	binomial variance of site occupancy P = M(R/M)[(M - R)/M]	dispersion <i>p</i> : variance of abundance among sites	species
column total	richness S: number of species at site	density s: number of individuals at site over all species	site
variance of column totals	variance of richness among sites $P_M$	variance of density among sites: overall dispersion $p_M$	set of M sites
column variance	binomial variance of species presence Q = N(S/N)[(N - S)/N]	evenness $q$ as measure of taxonomic heterogeneity	site
row × column interaction	variance of composition $T$ : environmental variance of $Q$ , specific variance of $R$	variance of relative abundance $t$ : environmental variance of $q$ , specific variance of $r$	set of N species and M sites

يرە

interactions may generate linear relationships (Loreau 2000) and because the relationship is sensitive to the spatial scale of surveys. Thus, an asymptotic local-regional plot may become linear when the area of different regions is taken into account (Shurin et al. 2000), and plots tend in general to become more nearly linear when the local area surveyed is large relative to the overall size of the region from which its species are drawn (Hillebrand & Blenckner 2002; Koleff & Gaston 2002).

In the NCM, local diversity for a given number of individuals is a power-law function of the immigration rate per species per cycle, m (figure 1). At high values of m, the power law has a slope z that approaches unity and thus indicates strict proportionality between regional and local diversity. As m decreases, z decreases, so that arithmetic axes would indicate an increasing degree of local community saturation. At low values of m, the power law breaks down and even the log-log plot shows saturation. These patterns are nevertheless caused by variation in immigration rate m alone; the community is equally (and completely) saturated in all cases. The local diversity of sedges, which have poor dispersal ability, is S = 46 species of a total regional pool of ca. N = 220 species (see Marie-Victorin 1964); a parallel survey of the more effectively dispersed ferns gave S = 38 and N = 75.

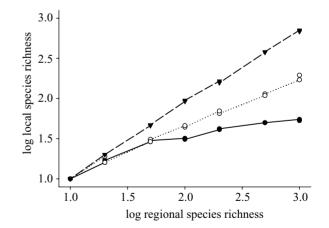


Figure 1. Local and regional processes. Relationship between species richness in a local community of 10 000 individuals and species richness of an external pool at three levels of immigration. Triangles, m = 0.1; open circles, m = 0.01; filled circles, m = 0.001.

#### (e) Distribution of abundance

The frequency distribution of abundance among species is a fundamental attribute of communities, because sampling from this distribution determines the level of species diversity. The statistical characterization of this distribution is a classical topic in theoretical ecology (Fisher et al. 1943; Preston 1948; May 1975), but does not in itself provide an understanding of the dynamic principles that cause abundance to vary. An alternative logical approach has been to show that if species are distributed in a selfsimilar manner, such that the fraction of species present in area A that also occur in area A/2 is independent of A, the distribution of abundance resembles the skewed lognormal characteristic of many communities (Harte et al. 1999). Again, this does not by itself identify the underlying ecological processes. Variation in abundance of plant species has often been linked to ecological variables such as seed size (e.g. Guo et al. 2000; Partel et al. 2001). Rarity, in particular, has been attributed to poor dispersal, habitat specialization, intolerance of disturbance and a variety of other ecological attributes (see Gaston 1994a; Bruno 2002). It has been firmly established, however, that the NCM will readily generate distributions of abundance resembling those of real communities (Hubbell 1995, 2001; Bell 2000).

The dominance-diversity curve is a useful way of representing the distribution of abundance over species when relatively few species have been surveyed (Whittaker 1965). The curve for 42 species of Carex in the 49 16 ha blocks at Mont St-Hilaire is shown in figure 2a. For comparison with the survey data, I used a community of 50 species on a  $7 \times 7$  grid with a low rate of dispersal (u = 0.01) between cells. The number of species and sites approximates those in the survey (including the possibility that a few rare species were present but not found, as was the case); a reasonable value for the dispersal rate, which is unknown, was contrived as follows. Imagine a narrow strip of width d around the margin of a cell of linear dimension L, such that offspring produced in this strip are just as likely to disperse to the neighbouring cell as they are to remain where they were born. The width of this strip is thus d = uL/2, so that for a 16 ha cell and a dispersal rate of 0.01 we have d = 2 m. Thus, u = 0.01 is roughly equivalent to an average dispersal distance of a few metres, which seems reasonable for seeds dispersed by gravity, ants and local water flows. Two realizations of this model yielded dominance-diversity curves resembling those of the survey (figure 2b).

Quantitative analyses show how the dominance–diversity curves of various communities can be fitted with great precision by the NCM (Hubbell 2001, ch. 5). Nonetheless, McGill (2003) has argued that although the NCM often explains 99% of variance in abundance, the empirical lognormal explains even more, and the NCM should for this reason be rejected as a mechanistic interpretation of the data. A statistical distribution is not a mechanistic hypothesis, however. The normal distribution often provides a very close fit to the frequency distribution of morphological attributes of individuals, such as size, but is not for this reason to be preferred to an interpretation in terms of the underlying genetic and environmental causes of variation.

#### (f) Nestedness

A community is nested insofar as the composition of a site with fewer species is a proper subset of sites with more; many different kinds of organism have strongly nested species assemblages and a variety of explanations

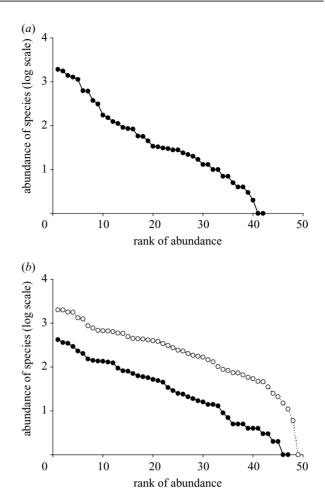


Figure 2. Dominance-diversity curves. Log abundance is plotted against rank of abundance. (a) Carex survey. Number of records of each of 42 species from 4144.25 ha cells at Mont St-Hilaire. (b) NCM. Two runs using a pool of 50 species and parameters as in figures 3-6; the upper curve has a limit of 500 individuals per cell, the lower curve of 100 individuals per cell.

have been proposed in terms of the ecological attributes of the species or sites involved (e.g. terrestrial invertebrates (Sfenthourakis et al. 1999), marine fish (McLain & Pratt 1999), birds (Calme & Desrochers 1999), mammals (Kelt et al. 1999), plants (Honnay et al. 1999; Butaye et al. 2001), and herpetofauna (Hecnar et al. 2002)). The importance of nestedness as a measure of community structure was emphasized by Worthen (1996). The degree of nestedness can be expressed as the number of 'gaps' in a species × sites matrix sorted by sites from the most to the least species-rich and by species from that with the largest to that with the smallest range (Patterson & Altmar 1986). A gap is the absence of a species from a site where it would be expected to occur if communities were perfectly nested. The Carex survey has 256 gaps, a fraction 0.127 of all available cells; the two NCMs have 379 (0.154) and 316 (0.129) gaps respectively. The degree of nestedness of NCM matrices increases with the rate of local dispersal, and is thus rather modest in these instances. For comparison, mammals from mountain ranges in the American southwest had a fraction 0.163 of gaps (Patterson & Altmar 1986), and Darwin's finches on the Galapagos islands a value of 0.295 (Worthen 1996). More sophisticated parameters of nestedness have been

devised (Brualdi & Sanderson 1999), along with tests to characterize the degree of nestedness found in randomized datasets (Cook & Quinn 1998), but these do not upset the conclusion that NCMs lead to about the same degree of nestedness as that found in surveys.

# 3. RELATIONSHIPS AMONG SURVEY VARIABLES

The direct effects of scale are obvious and well known; the number of species recorded, for example, will depend on the number of individuals sampled. There are indirect effects that are less obvious. I shall express these effects as power laws. These are relationships of the form  $y = ax^z$ , which can be linearized as log  $y = z \log x + a$ . The slope (exponent) z is independent of the units of measurement and thus directly comparable across surveys. To distinguish different power laws, the regression parameters referring to the relationship between y and x will be written z[y,x] and a[y,x]. There are two sets of power laws, one describing the variation of abundance among species and the other describing the variation of diversity among sites, which I discuss in that order.

#### (a) Power laws relating survey variables

Two very simple relationships can be described immediately. The first relates binary to quantitative scores of abundance (figure 3a): the number of sites from which a species is recorded increases with the number of individuals found,  $\log R = z[R,r]\log r + a[R,r]$ . This is an example of the range-abundance relationship, in which the abundance of a species is the total number of individuals present in the survey area (see Gaston 1996). The second relates variances to sums. The variance of abundance of a species among sites (dispersion) increases with the number of individuals found (figure 3c): log  $p = z[p,r]\log r + a[p,r]$ . The regression of log p on log r is akin to Taylor's power law (Taylor 1961), which has been used extensively in animal ecology to describe the degree of dispersion of individuals among populations. The NCM clearly generates very similar relationships between range (figure 3b) or dispersion (figure 3d) and abundance. The main difference for these examples is that range tends to reach an asymptote in the survey data, very abundant species being present in every quadrat; the NCM generates this outcome at greater values of local dispersal.

These two laws have transparent interpretations. However, they lead to three sets of derivative power laws obtained by chaining regression equations. These laws are more interesting, and include several well known ecological generalizations. I emphasize that the derivative power laws are algebraic, not statistical, results, and therefore differ from empirical relationships in the presence of error variance. The predicted values of their parameters can be expected to be approximately correct only when all their component regressions are well fitted. Nevertheless, the two basic power laws are sufficient to entail the existence, the direction and the approximate magnitude of these derivative laws.

The first law relates sums to means (figure 4*a*): the range of a species increases with mean local abundance: log  $R = z[R,r/R]\log(r/R) + a[R,r/R]$ . The expected value of the regression coefficient is z[R,r/R] = z[R,r]/(1 - z[R,r]),

although in practice the relationship saturates for species that are found nearly everywhere. This is an instance of the range–abundance relationship in which the abundance of a species is the mean number of individuals per cell, for those cells in which the species occurs (Sutherland & Baillie 1993; Gregory & Blackburn 1995). Gaston & Lawton (1990) suggest that the relationship is likely to be strongly positive when the reference locality at which abundance is estimated is representative of the entire area over which range is estimated; in these conditions, R is strongly correlated with r and the range–abundance relationship follows. Self-similarity leads to power-law range–abundance relationships with realistic slopes (Harte *et al.* 2001). The NCM generates a similar relationship (figure 4b).

The second law relates the variance of frequencies to sums (figure 4c): the relationship between the dispersion of frequencies among sites and the overall abundance of a species is given by  $\log \tilde{p} = z[\tilde{p},r]\log r + a[\tilde{p},r]$ , whose expected value is  $z[\tilde{p},r] = z[p,r] - 2$ . A value of -1 is expected if individuals were allocated at random to sites. The *Carex* data give a lower estimate of  $z[\tilde{p},r] = -0.765$ , indicating aggregation. The NCM generates a similar well-fitted power law with a negative exponent (figure 4d). The value of z(-0.412, -0.447) in two replicates) is somewhat lower, indicating a higher degree of aggregation; it could be tuned upwards by increasing the local disperal rate.

The third law relates the variance of frequencies and binary sums: the relationship between dispersion and range is expected to be  $\log \tilde{p} = z[\tilde{p},R]\log R + a[\tilde{p},R]$ , where  $z[\tilde{p},R] = z[\tilde{p},r]/z[R,r]$ . If  $z[\tilde{p},R] < 0$  then species with low overall range or abundance tend to be patchily distributed. This pattern was displayed both by the survey data (figure 4*e*) and by the NCM (figure 4*f*).

A completely parallel set of power laws can be developed for the variation of diversity among sites. These proceed from the basic power laws describing how species richness S (figure 5a) and the variance of abundance q(figure 5c) vary with the number of individuals sampled to the three derived power laws (figure 6a,c,e). Again, the NCM generates similar patterns (figures 5b, d and 6b, d, f), although species richness is only poorly correlated with abundance in the survey and even more poorly in the model. The most interesting relationship is that between species richness and evenness, the two most commonly used measures of diversity. Both Cook & Graham (1996) and Stirling & Wilsey (2001) found that species richness and evenness are usually negatively correlated in a variety of communities, and this is also true for the Carex data (figure 6e). A very similar negative relationship is generated by the NCMs, however (figure 6f). A more extended discussion of the relationship between richness and evenness in NCMs can be found in Bell (2000).

# 4. SPATIAL STRUCTURE

#### (a) The species-distance rule

If two sites separately support  $S_j$  and  $S_k$  species their joint diversity  $S_{jk}$  can be partitioned as follows:

$$S_{jk} = S_j + S_k - S_j S_k / N - (N-1) Cov(X_{ij}, X_{ik}),$$

where  $Cov(X_{ij}, X_{ik})$  is the binary genetic (specific) covari-

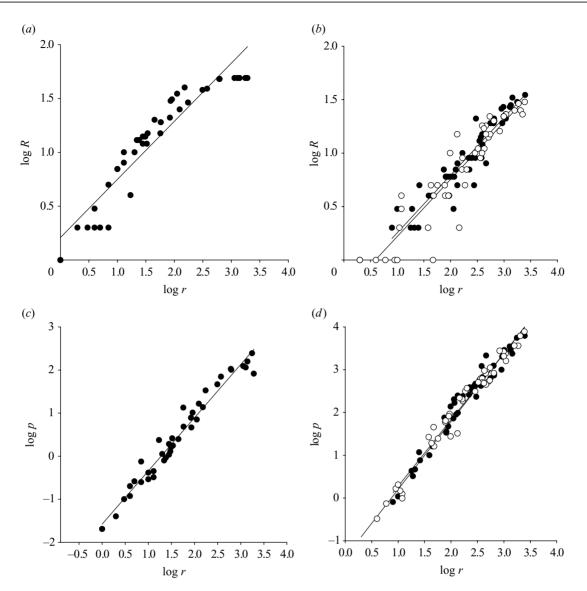


Figure 3. Basic power laws for range and dispersion of species. (a) and (c) refer to the Carex survey at Mont St-Hilaire described in the text; and (b) and (d) show how the output of two replicate runs of a NCM with: M = 49 sites (cells), regional pool N = 50 species, capacity K = 500 individuals per cell, probability of birth b = 0.505 per individual per cycle, probability of death d = 0.5 per individual per cycle, probability of immigration m = 0.005 per species per marginal cell per cycle, and probability of dispersal u = 0.01 for each newborn.

ance over all N species in the survey. If  $n_{11}$  denotes the number of cases in which both species are present,  $X_{ii} = 1$  and  $X_{ik} = 1$ , and so forth for all four combinations of presence and absence, then  $Cov(X_{ij}, X_{ik}) = (n_{11}n_{00} - n_{11}n_{00})$  $n_{10}n_{01})/N(N-1)$ . In a heterogeneous landscape, the genetic covariance will tend to fall with the distance between sites, because distant sites are more likely to provide different conditions of growth and thereby select for different combinations of species. Consequently, the combined species richness of sites will increase with distance, a pattern sometimes called 'turnover' (Gauch 1973; Whittaker 1975). Turnover is often expressed categorically as the difference between 'alpha' (within-site) and 'beta' (among-site) diversity (Whittaker 1975). The rate at which correlation decays with distance, or equivalently the balance of alpha and beta diversity, may reflect the relative importance of local and regional processes. Nekola & White (1999) have provided a detailed account of turnover in plant communities of eastern North America, attributing the decay of correlation with distance either to

Proc. R. Soc. Lond. B (2003)

local adaptation to an heterogeneous environment or to limited local dispersal.

In the *Carex* survey at Mont St-Hilaire, the covariance falls and the combined species richness increases, as expected under local adaptation (figure 7*a*). A similar pattern emerges from the NCM, however, where it is generated solely by local dispersal (Figure 7*b*). Nekola & White (1999) found that the rate of decay of correlation was greater for species of intermediate abundance and was also greater in fragmented landscapes. Both results are consistent with the NCM. They also found that the rate of decay varied among groups of organisms in a way suggesting that it was directly related to the rate of local dispersal; this is also predicted by the NCM, except that the model would then strictly apply only within groups whose members have similar powers of dispersal.

#### (b) The species-area rule

The increase in overall species richness with the extent of the survey is among the most familiar generalizations

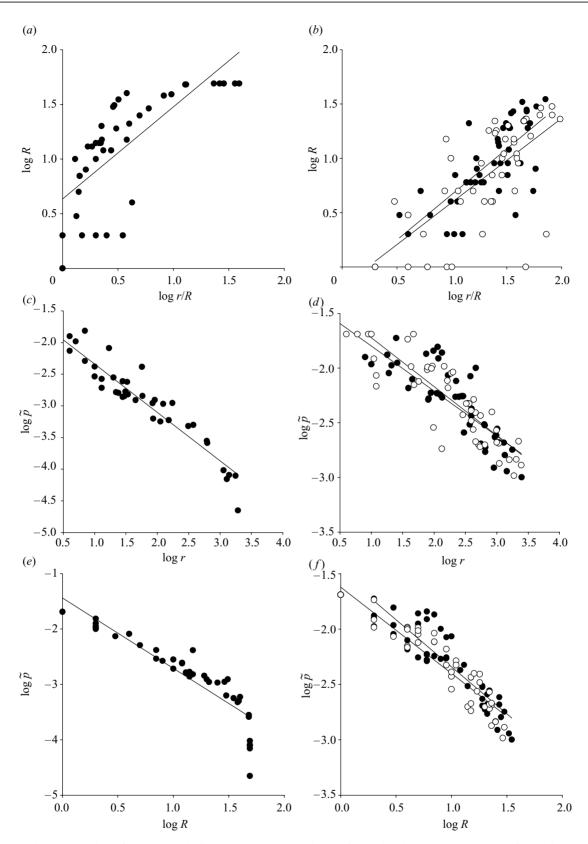


Figure 4. Derived power laws for range and dispersion. (a), (c) and (e) refer to the *Carex* survey at Mont St-Hilaire described in the text; and (b), (d) and (f) show the output of two replicate runs of the NCM (parameters as in figure 3).

in ecology and is the subject of a large literature (reviewed by Rosenzweig 1995). To the extent that the relationship depends on sampling, its form is determined by the distribution of abundance among species. If diversity maps are self-similar then it can be shown that the species-area relationship is expected to be a power law of the form  $S = \text{constant} \times \text{area}^z$  (Harte *et al.* 1999), although the success of this fractal analysis in predicting the relationship precisely has been questioned (Ostling *et al.* 2000; Plotkin *et al.* 2000). Because the NCM accurately predicts the

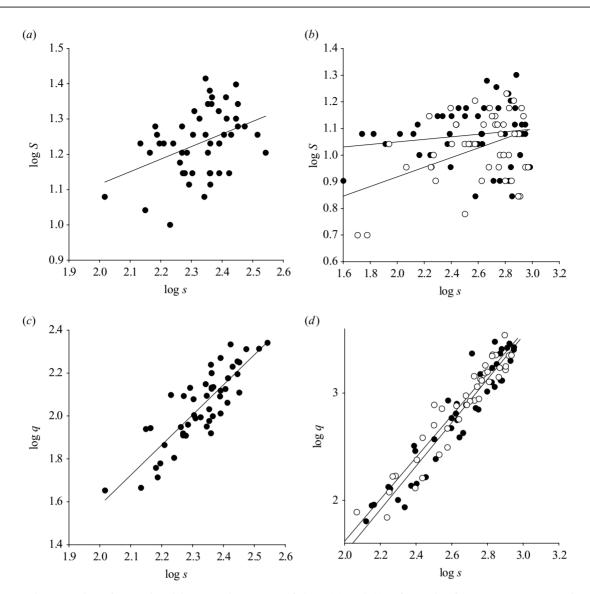


Figure 5. Basic power laws for species richness and evenness of sites. (a) and (c) refer to the Carex survey at Mont St-Hilaire described in the text; and (b) and (d) show the output of two replicate runs of the NCM (parameters as in figure 3).

form of the abundance distribution it can also be expected to predict the form of the species-area relationship from mechanistic principles (Borda-de-Agua *et al.* 2002).

The power law relating species richness to the area of nested blocks in the *Carex* survey is well fitted  $(r^2 = 0.992)$ , although detectably curvilinear, and has z = 0.360 (figure 7c). An equally well-fitted power law species–area relationship is readily generated by the NCM (Bell 2001; Hubbell 2001), and can readily be made to bracket the survey result by tuning capacity from 50 individuals (z = 0.427) to 100 individuals (z = 0.283) per cell (figure 7d).

# 5. DISCUSSION

Biological surveys often discover high levels of diversity in natural communities; in the example I have used, about 50 species of a single genus of sedges have been found in 1000 ha of old-growth forest. One theory of how this diversity is maintained is that each species is well adapted to some combination of environmental factor states, representing a small fraction of the available range of environmental variation, so that it is able to exclude competitors from sites where this combination occurs. The term 'factor' can be interpreted very broadly to include soil pH, particular species of pathogen, time since disturbance or any of many other possibilities; in any case, environmental heterogeneity gives rise to divergent natural selection, and each kind of site will usually be occupied by the few species best adapted to it. This adaptationist interpretation of diversity has a serious flaw. The coexistence of 50 species, even when the survey is confined to a single genus in a single forest fragment, would require a correspondingly high degree of specialization, the 'paradox of the plankton' noted long ago by Hutchinson (1961). The most obvious symptom of this fine-scale niche differentiation would be the inability of most species to grow in most sites. This would impose a severe dispersal load on the population, because very few propagules would reach their preferred sites, and many sites would be empty. This is not in fact observed: most species can be transplanted successfully to most sites (Bell et al. 2000), and more generally this degree of ecological intolerance is rarely observed. There is therefore a need for an alternative general theory of species diversity, which the neutral theory provides. It suffers from an equally irksome difficulty: field naturalists

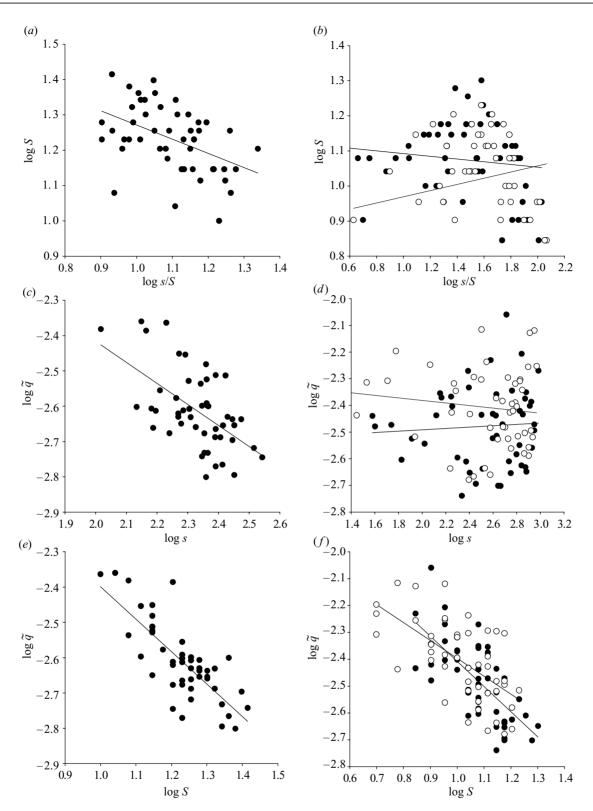


Figure 6. Derived power laws for species richness and evenness. (a), (c) and (e) refer to the Carex survey at Mont St-Hilaire described in the text; and (b), (d) and (f) show the output of two replicate runs of the NCM (parameters as in figure 3).

know it to be wrong because many species are consistently found at characteristic kinds of site.

It might be hoped that analysing the patterns that emerge from systematic biological surveys would resolve this dilemma by providing decisive evidence of the operation of ecological mechanisms such as niche differentiation. This has indeed been the goal, implicit or explicit, of the very large literature that is devoted to identifying and analysing these patterns. Ecologists have often been warned that processes cannot necessarily be inferred reliably from patterns (Lawton 1999), and the main point of this article is to show that these warnings are fully justified. All of the strongly marked patterns that emerge from survey data emerge in the same form from simple neutral models, and therefore do not help us to distinguish between adaptationist and neutralist interpretations of species diversity.

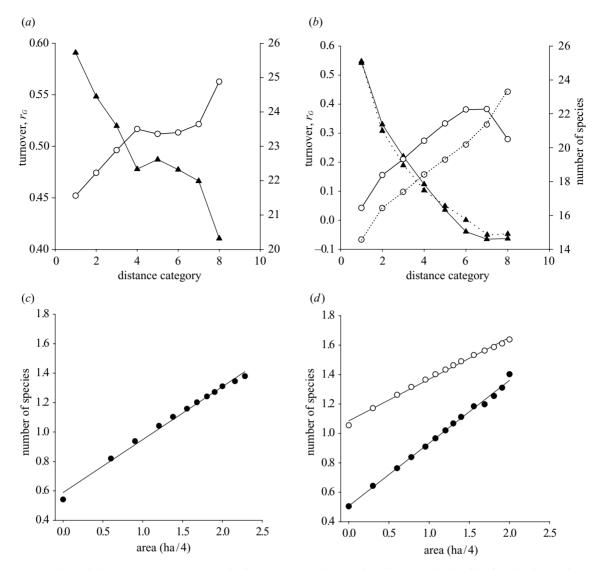


Figure 7. General spatial patterns (parameters as in figure 3). (a) The species-distance relationship for the Carex dataset. Genetic (specific) correlation falls as the distance between sites increases (filled triangles), so the combined species richness of the two sites increases (open circles). (b) The species-distance relationship for the NCM. Two replicate runs with the same parameter values are shown. (c) The species-area relationship for nested blocks of the *Carex* dataset. The exponent of the power law is z = 0.360. (d) The species-area relationship for nested blocks of a NCM with 3600 cells; parameters as before except u = 0.1 to compensate for smaller cell size. The lower curve (z = 0.427) refers to a community with twice as many individuals per cell as the upper curve (z = 0.283).

One response to this conclusion is to point to species that are clearly specialized to a narrow range of conditions and are unable to sustain themselves elsewhere. Many examples could be cited, but unless the list can be expanded to include most of the species in the local community it does not resolve the issue at hand: what proportion of local diversity is maintained by local selection? Alternatively, more subtle or less intense processes of selection might favour different species in different kinds of site, even though most species are able to grow in most sites in the absence of competitors. The question then becomes whether survey data will enable us to distinguish between weak selection and no selection. For the simple patterns discussed in this article, I think it is very unlikely that any such attempt would be successful. More informative patterns might exist, but so far they have been elusive. Clark & McLachlan (2003) point out, for example, that although the variance of independent stochastic processes is expected to increase through time,

to show any consistent tendency to increase since the completion of post-glacial recolonization. But sites are not independent, because of local dispersal; and quite low rates of dispersal, of the order of the reciprocal of the local population size, are sufficient to prevent any consistent trend in variance among sites. From the survey itself, the co-distribution of species could be analysed to detect the expected surplus of combinations of species that were found at the same site much more often, or much less often, than expected. Furthermore, if environmental data were also available, the relationship between diversity and environmental factor states could be investigated directly to identify the ecological basis of variation in diversity. Analyses of this sort have been completed for the Mont St-Hilaire survey, and will be published separately. Without wishing to anticipate their conclusions in detail, they have not left me more optimistic about using survey data

the variance of abundance of tree taxa (as judged from pollen records) among sites in southern Ontario has failed

to resolve decisively the two leading interpretations of diversity.

The very similar issue of genetic diversity within populations was debated at great length by population geneticists some 30 years ago. Community ecologists might refer to niche differentiation arising through competition, whereas population geneticists would refer to adaptation caused by natural selection; I have treated these processes as being interchangeable to underline the parallel between neutral models of communities and populations. A masterly half-time report was presented by Richard Lewontin, who after an extended summary of the available categories of evidence rather forlornly asked, 'how can such a rich theoretical structure as population genetics fail so completely to cope with the body of fact?' (Lewontin 1974, p. 267). To answer his own question, he pointed first to the empirical insufficiency of the theory, because its predictions involve combinations of parameters such as Nu, where N is a population size and u a mutation rate. N is a very large number, u is a very small number, and both are extremely difficult to estimate with precision; because guesses of Nu may in consequence range over orders of magnitude, it is rarely possible for the theory to exclude any observation. In community ecology the corresponding quantity is Nv, where v is a speciation rate (Hubbell version of the neutral model), or Nm, where m is a dispersal rate (this version). The same difficulty applies. The study of species diversity has thus arrived at a crossroads: the NCM provides a simple and powerful general theory, but it has not been possible so far to distinguish it clearly from adaptationist interpretations using survey data. In population genetics this difficulty never was resolved by the contemplation of pattern or the compilation of examples; instead, it was speedily settled in the late 1980s, when it became possible to distinguish between coding and noncoding nucleotide substitutions. It may require a technological advance of similar magnitude to perform the same service for ecology.

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and by la Fondation pour les Chercheurs et à l'Aide de la Recherche du Québec.

#### REFERENCES

- Bell, G. 2000 The distribution of abundance in neutral communities. Am. Nat. 155, 606–617.
- Bell, G. 2001 Neutral macroecology. Science 293, 2413-2418.
- Bell, G., Lechowicz, M. J. & Waterway, M. 2000 Environmental heterogeneity and species diversity of forest sedges. J. Ecol. 88, 67–87.
- Borda-de-Agua, L., Hubbell, S. P. & McAllister, M. 2002 Species-area curves, diversity indices, and species abundance distributions: a multifractal analysis. *Am. Nat* 159, 138–155.
- Brualdi, R. A. & Sanderson, J. G. 1999 Nested species subsets, gaps and discrepancy. *Oecologia* **119**, 256–264.
- Bruno, J. F. 2002 Causes of landscape-scale rarity in cobblebeach plant communities. *Ecology* 83, 2304–2314.
- Butaye, J., Jacquemyn, H. & Hermy, M. 2001 Differential colonization causing non-random forest plant community structure in a fragmented agricultural landscape. *Ecography* 24, 369–380.

- Calme, S. & Desrochers, A. 1999 Nested bird and micro-habitat assemblages in a peatland archipelago. *Oecologia* **118**, 361–370.
- Clark, J. S. & McLachlan, J. S. 2003 Stability of forest biodiversity. *Nature* **423**, 635–638.
- Cody, M. L. & Diamond, J. M. (eds) 1975 *Ecology and evolution of communities.* Cambridge, MA: Harvard University Press.
- Cook, L. M. & Graham, C. S. 1996 Evenness and species number in some moth populations. *Biol. J. Linn. Soc.* 58, 75–84.
- Cook, R. R. & Quinn, J. F. 1998 An evaluation of randomization models for nested species subsets analysis. *Oecologia* 113, 584–592.
- Cornell, H. V. & Lawton, J. H. 1992 Species interactions, local and regional processes, and limits to richness of ecological communities: a theoretical perspective. *J. Anim. Ecol.* 61, 1–12.
- Diamond, J. M. & Case, T. J. (eds) 1986 *Community ecology*. New York: Harper & Row.
- Enquist, B. J., Sanderson, J. & Weiser, M. D. 2002 Modelling macroscopic patterns in ecology. *Science* 295, 1835–1837.
- Fisher, R. A., Corbet, A. S. & Williams, C. 1943 The relationship between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12, 42–58.
- Frontier, S. 1985 Diversity and structure in aquatic ecosystems. Oceanogr. Mar. Biol. A. Rev. 23, 253-312.
- Gaston, K. J. 1994a Rarity. London: Chapman & Hall.
- Gaston, K. J. 1994b Measuring geographic range sizes. Ecography 17, 198–205.
- Gaston, K. J. 1996 The multiple forms of the interspecific abundance-distribution relationship. *Oikos* 76, 211–220.
- Gaston, K. J. & Lawton, J. H. 1990 Effects of scale and habitat on the relationship between regional distribution and local abundance. *Oikos* 58, 329–335.
- Gauch, H. G. 1973 The relationship between sample similarity and ecological distance. *Ecology* 54, 618–622.
- Gee, J. H. R. & Giller, P. S. (eds) 1987 Organization of communities past and present. Oxford: Blackwell.
- Gregory, R. D. & Blackburn, T. M. 1995 Abundance and body size in British birds: reconciling regional and ecological densities. *Oikos* 72, 151–154.
- Guo, Q., Brown, J. H., Valone, T. J. & Kachman, S. D. 2000 Constraints of seed size on plant distribution and abundance. *Ecology* 81, 2149–2155.
- Harper, J. L. & Hawksworth, D. L. 1994 Biodiversity: measurement and estimation (preface). *Phil. Trans. R. Soc. Lond.* B **345**, 5–12.
- Harte, J., Kinzig, A. & Green, J. 1999 Self-similarity in the distribution and abundance of species. *Science* 284, 334–336.
- Harte, J., Blackburn, T. & Ostling, A. 2001 Self-similarity and the relationship between abundance and range size. *Am. Nat.* 157, 374–386.
- Hecnar, S. J., Casper, G. S., Russell, R. W., Hecnar, D. R. & Robinson, J. N. 2002 Nested species assemblages of amphibians and reptiles on islands in the Laurentian Great Lakes. *J. Biogeogr.* 29, 475–489.
- Hillebrand, H. & Blenckner, T. 2002 Regional and local impact on species diversity: from pattern to processes. *Oecol*ogia 132, 479–491.
- Honnay, O., Hermy, M. & Coppin, P. 1999 Nested plant communities in deciduous forest fragments: species relaxation or nested habitats? *Oikos* 84, 119–129.
- Hubbell, S. P. 1995 Toward a theory of biodiversity and biogeography on continuous landscapes. In *Preparing for global change: a mid-western perspective* (ed. G. R. Carmichael, G. E. Folk & J. E. Schnoor). Amsterdam: Academic.

- Hubbell, S. P. 2001 *The unified neutral theory of biodiversity and biogeography.* Princeton University Press.
- Hutchinson, G. E. 1961 The paradox of the plankton. Am. Nat. 95, 137-145.
- Kelt, D., Rogovin, K., Shenbrot, G. & Brown, J. H. 1999 Patterns in the structure of Asian and North American desert small mammal communities. J. Biogeogr. 26, 825–841.
- Kempton, R. A. 1979 The structure of species abundance and measurement of diversity. *Biometrics* 35, 307–321.
- Kimura, M. 1983 *The neutral theory of molecular evolution*. Cambridge University Press.
- Koleff, P. & Gaston, K. J. 2002 The relationships between local and regional species richness and spatial turnover. *Global Ecol. Biogeogr. Lett.* 11, 363–375.
- Lande, R. 1996 Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76, 5–13.
- Lawton, J. H. 1999 Are there general laws in ecology? *Oikos* 84, 177-192.
- Levins, R. 1969 Some demographic and genetic consequences of environmental heterogeneity in biological control. *Bull. Entomol. Soc. Am.* 15, 237–240.
- Lewontin, R. C. 1974 The genetic basis of evolutionary change. New York: Columbia University Press.
- Loreau, M. 2000 Are communities saturated? On the relationship between alpha, beta and gamma diversity *Ecol. Lett.* **3**, 73–76.
- Macarthur, R. H. & Wilson, E. O. 1967 The theory of island biogeography. Princeton University Press.
- McGill, B. J. 2003 A test of the unified neutral theory of biodiversity. *Nature* 422, 881–885.
- McLain, D. K. & Pratt, A. E. 1999 Nestedness of coral reef fish across a set of fringing reefs. *Oikos* 85, 53–67.
- Marie-Victorin, F. 1964 *Flore laurentienne*. Les Presses de l'Université de Montréal.
- May, R. M. 1975 Patterns of species abundance and diversity. In *Ecology and evolution of communities* (ed. M. L. Cody & J. M. Diamond), pp. 81–120. Cambridge, MA: Harvard University Press.
- Nekola, J. C. & White, P. S. 1999 The distance decay of similarity in biogeography and ecology. J. Biogeogr. 26, 867–878.
- Ostling, J., Harte, J., Green, J. & Condit, R. 2000 Selfsimilarity and clustering in the spatial distribution of species. *Science* **290**, 671.
- Partel, M., Moora, M. & Zobel, M. 2001 Variation in species richness within and between calcareous (alvar) grassland stands: the role of core and satellite species. *Pl. Ecol.* 157, 205–213.
- Patterson, B. D. & Altmar, W. 1986 Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biol. J. Linn. Soc.* 28, 65–82.

- Plotkin, J. B. (and 12 others) 2000 Predicting species diversity in tropical forests. *Proc. Natl Acad. Sci. USA* 97, 10 850– 10 854.
- Preston, F. W. 1948 The commonness, and rarity, of species. *Ecology* **29**, 254–283.
- Ricklefs, R. E. 1987 Community diversity: relative roles of local and regional processes. *Science* 235, 167–171.
- Rosenzweig, M. L. 1995 Species diversity in space and time. Cambridge University Press.
- Sfenthourakis, S., Giokas, S. & Mylonas, M. 1999 Testing for nestedness in the terrestrial isopods and snails of Kyklades islands (Aegean archipelago, Greece). *Ecography* 22, 384– 395.
- Shurin, J. B., Havel, J. L., Leibold, M. A. & Pinel-Alloul, B. 2000 Local and regional zooplankton species richness: a scale-independent test for saturation. *Ecology* 81, 3062– 3073.
- Simberloff, D. & Connor, E. F. 1979 Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. In *Contemporary quantitative* ecology and related ecometrics (ed. G. P. Patil & M. L. Rosenzweig), pp. 123–128. Fairland, MD: International Cooperative Publishing House.
- Smith, B. & Wilson, J. B. 1996 A consumer's guide to evenness indices. *Oikos* 76, 70–82.
- Srivastava, D. S. 1999 Using local-regional richness plots to test for species saturation: pitfalls and potentials. J. Anim. Ecol. 68, 1–16.
- Stirling, G. & Wilsey, B. 2001 Empirical relationships between species richness, evenness, and proportional diversity. *Am. Nat.* 158, 286–299.
- Sutherland, W. J. & Baillie, S. R. 1993 Patterns in the distribution, abundance and variation of bird populations. *Ibis* 135, 209–210.
- Taylor, L. R. 1961 Aggregation, variance and the mean. *Nature* **189**, 732–735.
- Washington, H. G. 1984 Diversity, biotic and similarity indices. Water Res. 18, 653–694.
- Weiher, E. & Keddy, P. 1999 Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press.
- Whitfield, J. 2002 Neutrality versus the niche. *Nature* 417, 480–481.
- Whittaker, R. H. 1965 Dominance and diversity in land plant communities. *Science* 147, 250–260.
- Whittaker, R. H. 1975 Communities and ecosystems. New York: Macmillan.
- Worthen, W. B. 1996 Community composition and nestedsubset analyses: basic descriptors for community ecology. *Oikos* 76, 417–426.