THE CO-DISTRIBUTION OF SPECIES IN RELATION TO THE NEUTRAL THEORY OF COMMUNITY ECOLOGY

GRAHAM BELL¹

Redpath Museum and Biology Department, McGill University, Montreal, Québec, H3A 2K6, Canada

Abstract. Patterns of species co-distribution have often been used to infer the operation of processes such as competition and selection among species. Ecological specialization should create unexpectedly strong species associations, reflected by high positive or negative correlations between species that are adapted to similar or different kinds of site.

Two systematic surveys of plant communities do indeed show stronger associations than are expected in randomly assembled communities. Neutral community models, however, also predict nonrandom patterns of co-distribution. Local selection in heterogeneous landscapes, where species are sorted into the sites to which they are best adapted, will produce even stronger associations, provided that different kinds of site are correlated in space, forming repeated combinations or "habitats." In practice, however, this effect is readily detectable only for intense selection in coarse-grained landscapes. In contrast, classification of species or sites into prior categories consistently produces positive associations, and thereby demonstrates the existence of non-neutral processes structuring communities. It is concluded that the bulk properties of communities, such as overall diversity, may be strongly affected by local dispersal and stochastic drift and, for this reason, are adequately represented by neutral models in many cases, despite the operation of systematic processes of local selection. One corollary of this interpretation is that these bulk properties are unlikely to provide useful information about community processes.

Key words: abundance; adaptation; assembly rules; co-occurrence; dispersal; distribution; diversity; neutral model; range.

INTRODUCTION

Most species are unevenly distributed over the landscape at any given spatial scale, and communities differ in composition from site to site. These familiar facts can be given two rival interpretations, the two poles of community ecology theory. The first is that each species occupies only those sites to which it is well adapted, and from which it is able to exclude competitors. The community appears to be well engineered, so to speak, in the sense that the loss of any of its specialized components is likely to reduce the productivity or the stability of the system as a whole. This view underlies the phytosociological approach to communities, which interprets the landscape as a mosaic of distinct and homogeneous habitats each occupied by a characteristic set of species (Braun-Blanquet 1932), and it implies the existence of assembly rules that govern the composition of the community that will develop at a particular site (Diamond 1975). The second interpretation is that most species are able to grow at most sites, so that community composition is determined largely by the accidents of dispersal, and local diversity is strongly influenced by the composition of the regional species pool. An extreme view is that all eco-

Manuscript received 28 June 2004; revised 27 October 2004; accepted 29 October 2004; final version received 17 December 2004. Corresponding Editor: N. J. Gotelli.

¹ E-mail: Graham.bell@mcgill.ca

logically similar species (MacArthur and Wilson 1967) or all individuals belonging to ecologically similar species (Hubbell 1995, 2001, Bell 2000, 2001) have identical demographic attributes, so that community diversity and composition are determined entirely by immigration and dispersal. It follows that the loss of species will have no effect on higher-level community properties. This dichotomy between deterministic and stochastic, or machine-like and incoherent, worldviews is not peculiar to community ecology, but can be found in fields as disparate as population genetics and politics. In practice, most people might adopt an intermediate position, but their approach to the field is likely to be strongly influenced nonetheless by leaning more towards one view or the other.

It might be expected that theories so radically different would predict correspondingly different patterns of distribution and diversity. The patterns that emerge from biological surveys could then be used to evaluate ecological processes, and in particular the generality and intensity of competitive exclusion. The statistical analysis of co-distribution in plant communities was pioneered by Goodall (1953) and by Williams and Lambert (1959), who wrote that "vegetation . . . is a matrix of correlation coefficients." As the quotation suggests, much of this literature is a statistical exercise in the description of vegetation. The analysis of distribution and co-distribution, however, was also used in order to detect nonrandom patterns that must, it was felt, reflect

the determination of occurrence and performance by one or a few environmental factors (Greig-Smith 1961). The goal of such analyses was to show that all or most species in a community occupy different kinds of sites, each with characteristic conditions of growth. This is clearly true over very large areas where distributions are completely disjunct (coconut palms and black spruce, or a thousand other examples), but whether it remains true for much smaller areas, for species that grow intermingled, is less obvious. Gilbert and Lechowicz (2004) report a detailed survey of forest plants where they partition environmental from purely spatial variation, and conclude that environmental factors govern community composition at distances of 100-1000 m. Careful studies such as this provide strong evidence for the functional interpretation of communities, with the reservation that the relevant environmental factors must be specified in advance, and all included in the analysis. If a large number of factors are measured, and then the data set reduced by excluding those factors that fail some permissive criterion for significance, then strong and highly significant environmental associations are likely to appear even in random data (Freedman 1983). A more practical limitation is that biological surveys are, in any case, often conducted without measuring these factors, even supposing that they have been identified. Both difficulties can be surmounted, however, by analyzing patterns of species occurrence directly, without any reference to external variables, on the grounds that the species will themselves integrate all the relevant environmental factors (for differing opinions, see Wiens 1981, Gilpin and Diamond 1982, Wilson 1991). Surveys that record the presence or absence of species in a number of sites are readily available, and if they have been conducted carefully and systematically, can be used to evaluate associations among species. The purpose of this article is ask whether these associations, estimated from the best available systematic surveys, provide a signal that will enable us to distinguish neutral from functional interpretations of diversity.

METHODS

Landscape structure

A community consisting of many ecologically similar species inhabits a landscape, whose structure may affect species distributions. The landscape can be represented as an array of sites that are treated as being spatially distinct. Each site is capable of supporting no more than a maximum number of individuals, and is thus the locus of density regulation. Any number of contiguous sites collectively comprise a patch. The conditions of growth at any given site constitute one of a fixed range of discrete substrates. The relative fitness of species may vary among substrates, and the site is thus the locus of selection. Where there are several or many substrates, the landscape (as an array of physical units) is structured, and if this creates local selection, the environment (as an array of conditions of growth) can be said to be heterogeneous. The combination of substrates in a patch constitutes a habitat. This is a meaningful term only if the same few sets of substrates are consistently found close together, in which case the environment can be said to be correlated. Thus, "site" and "patch" are terms that denote spatial relationship alone, whereas "substrate" and "habitat" refer to the ecological characteristics of sites and patches, wherever they might happen to be. Finally, the grain of any survey imposes a lower spatial limit on the description of species distributions. In models, it is convenient to equate the site with the grain of the survey, so that there is no spatial or ecological variation within the site. This is unlikely to be true for any real survey, and it can never be ruled out that hidden patterns may exist among sites at spatial scales less than that of the grain of sampling. On the other hand, grain may readily be increased by aggregating neighboring sites to constitute a block. This cumbersome but necessary terminology is illustrated in Fig. 1.

Spatial community models

A spatially explicit neutral community model comprises a set of M sites, between which newborn individuals disperse with probability u. The basic neutral community model used here is the immigration-based model used in previous articles (Bell 2001, 2003); other versions are reviewed by Chave (2004). It can be modified to allow variation in relative fitness among sites. In the "habitat" model, each site or block of adjacent sites was allocated to one of a fixed number of distinct substrates. Individuals belonging to any given species had unit fitness (determined by the overall birth and death rates) in one of these substrates and fitness 1-sin all others, where s is a selection coefficient. In the "gradient" model, individuals belonging to any given species have unit fitness at some site, their fitness declining with distance away from this site in either direction, following a Normal function with standard deviation σ . A habitat model with s very small or a gradient model with σ very large will approach a neutral model.

Biological surveys

The output of the neutral community model (NCM) can be compared with the patterns displayed by systematic biological surveys. The ability of a survey to describe spatial patterns is constrained by its grain and extent. The ideal survey has small grain and large extent, but this is usually impracticable, and there is a very general tendency for grain to increase with extent. I have chosen two surveys of different grain and extent, recording the plant communities of very different landscapes, in which an area was systematically and completely searched by trained observers over a short period of time. The first was conducted in a large fragment



FIG. 1. Environmental structure, illustrating terms used in the text.

of old-growth forest at Mont St.-Hilaire, southern Quebec, ~1000 ha in extent, that was surveyed at 1-ha grain for ferns and at 0.25-ha grain for sedges (*Carex*), herbs, and trees (see Bell et al. 2001). The second survey was conducted throughout the county of Leicestershire, England, ~2000 km² in extent, surveyed for all plants at 4-km² grain (Primavesi and Evans 1988). This is a highly humanized landscape from which almost all of the original forest cover has long since been removed. The Leicestershire survey was chosen in part because it was supplemented by detailed independent surveys of ~100 sites, most a few hectares in extent, that were specially chosen to represent 12 distinct habitats.

Measures of co-distribution

The measure of site association that I shall use is the straightforward binary covariance of species among sites. The data are the presence or absence of the *i*th species at a given site, $X_i = 1$ or 0. For any pair of species over all *M* sites, both species occur at n_{11} sites, the first but not the second at n_{10} sites, the second but not the first at n_{01} sites, and neither occur in n_{00} sites. The binary covariance is then

$$\operatorname{Cov}(X_i, X_i) = (n_{11} n_{00} - n_{10} n_{01})/M(M - 1)$$

which can be expressed in standardized form as a correlation coefficient, in this case representing environmental correlation, r_{species} , by dividing by the geometric mean of the variances. This corresponds to the parametric correlation coefficient introduced as a measure of species association by Kershaw (1960), and to the "inverse" analysis of association by chi-square used by Williams and Lambert (1961). In a precisely parallel manner, one can compute the binary covariance of sites among species and standardize this as a genetic (or specific) correlation coefficient, r_{sites} (Bell 2003; cf. Qmode and R-mode analysis, Simberloff and Connor 1979). Local adaptation is expected to generate unexpectedly large positive and negative values of both correlations, such that the standard deviation of the frequency distribution of values of the correlation coefficient over all pairwise combinations of sites, or species, can be used to detect and to evaluate departures from randomness or from neutrality (Fig. 2).

Several other indexes related to the binary covariance and correlation coefficient have been proposed to express the degree of species co-distribution (analytical review by Gotelli 2000). Two of the most widely used are the *V* statistic of Pielou and Robson (1972; see Schluter 1984) and the checkerboard score *C* introduced by Stone and Roberts (1990). Pielou and Robson's *V* is the ratio of two variances: the variance of species richness *S* per site, Var(*S*), and the sum of the variances Var(X_i) of species occurrence over sites. The two are related by the following:



FIG. 2. Local selection is expected to increase the variance of species associations.

$$\operatorname{Var}(S) = \frac{\operatorname{species}}{\sum \operatorname{Var}(X_i)} + 2 \sum_{ij} \operatorname{Cov}(X_i, X_j).$$

The regression of V on r_{species} for random data has expected slope (N - 1) and intercept 1. Its expected value is thus 1, given that the net covariance is zero. The main limitation of this statistic is that it expresses only the degree to which species associations depart from a net covariance of zero; this might be misleading if some species pairs are positively and others negatively associated. The C score counts the number of cases over all pairs of sites in which two species occupy different sites. For any two species, this is $C_{ii} = (R_i - R_{ii})(R_i - R_{ii})$ R_{ij}), where R_i is the range of species *i* and R_{ij} the joint range of species *i* and *j*. Then $C = \sum \sum C_{ii}/P$, where *P* = S(S - 1)/2 is the number of pairwise comparisons among species. The C score is uncorrelated with $s(r_{\text{species}})$ in random data, but the two become rather strongly correlated when large positive and negative species associations are present. The main drawback of the C score is that its value depends on the number of sites surveyed. This can be mitigated by noting that $R_i = n_{10} + n_{11}, R_j = n_{01} + n_{11}$, and $R_{ij} = n_{11}$, from which it follows that

$$C = M(M - 1) \sum_{ij} \operatorname{Cov}(X_i, X_j)$$

under the constraint that joint absences are excluded. Thus, the modified index $C^* = C/M(M - 1)$ is equivalent to a binary covariance, under the constraint that joint absences are excluded, from which a correlation coefficient corresponding to r_{species} could be calculated if required.

RESULTS

Co-distribution in survey data is not random

These measures of co-distribution were calculated from the two surveys for several plant groups defined by relatedness, habit, phenology, or abundance. They were then compared with values gotten from randomized versions of the data sets (Table 1). In all cases, V > 1, showing that species are not distributed at random with respect to one another. The observed vale of $r_{\rm species}$ was slightly positive for all groups, in contrast to an expected value of zero. Consequently, the observed value of C was less than the value for randomized data, although this difference was smaller when the randomization was constrained by keeping fixed the number of records for each species. The variation of species associations, $s(r_{\text{species}})$, was greater in the survey data than in the randomized data, and was insensitive to the method of randomization. All of these differences were very large relative to the corresponding standard errors. In short, survey data show highly nonrandom patterns of co-distribution, for whatever reason.

Co-distribution in neutral models is not random

The output of a range of spatial NCMs can be compared with randomized data in the same way (Table 2). Net covariance was sometimes negative and sometimes positive, so that values of V were correspondingly less than or greater than 1, whereas values of C were correspondingly greater than or less than their expected values. The variation of co-distribution $s(r_{\text{species}})$ always exceeded that of randomized data. All differences are again very large relative to the corresponding standard errors. This shows that local dispersal alone can give rise to highly nonrandom patterns of co-distribution in neutral community models. Consequently, nonrandom patterns of co-distribution do not necessarily imply the operation of non-neutral processes such as competitive exclusion.

Co-distribution depends on selection intensity and landscape structure

For any particular set of parameters, local selection in a correlated landscape will create more extreme patterns of co-distribution than the corresponding neutral model. These parameters are unknown or poorly known in any particular case, however, and an arbitrary choice of parameters might not give a representative result. I have addressed this problem by first choosing an arbitrary but reasonable "kernel" of values for the set of demographic parameters that collectively specify the neutral model. For a particular realization of the model, a new parameter set is constructed from this kernel by increasing or decreasing the value of each parameter, and then choosing a new value from a uniform random distribution spanning a large range of possible values (see note to Table 3). Fifty independent realizations implemented in this fashion explored a very large area of parameter space, including the whole range of possible values for birth, death, and immigration rates, and a range spanning an order of magnitude for population size and local dispersal. They confirm that neutral models tend to develop nonrandom patterns of co-distribution (Table 3). Weak selection in a correlated envi-

	No	No	Grain	r _{spec}	ties	$s(r_{s_{I}})$	$s(r_{\rm species})$		V		C	
Group	species	sites	(ha)	Mean	SE	Mean	SE	Mean	SE	Mean	SE	
Mont StHilaire												
Ferns (survey) Random A	38	970	1	$0.0444 \\ -0.0003$	0.0003	$0.1021 \\ 0.0320$	0.0001	2.80 0.99	0.01	5753 18 294	12	
Random B Carex (survey)	42	4144	0.25	$0.0000 \\ 0.0176$	0.0004	$0.0317 \\ 0.0550$	0.0003	$\begin{array}{c} 1.01 \\ 1.77 \end{array}$	0.01	8782 38 915	14	
Random A Random B				$0.0001 \\ 0.0000$	$0.0001 \\ 0.0001$	0.0154 0.0157	$0.0001 \\ 0.0002$	$\begin{array}{c} 1.00 \\ 1.01 \end{array}$	$\begin{array}{c} 0.00 \\ 0.00 \end{array}$	99 305 44 557	22 43	
Spring herbs (survey)	42	4144	0.25	0.0273		0.1079		2.24		296 191		
Random A Random B Trees (survey)	41	4116	0.25	$0.0000 \\ -0.0002 \\ 0.0293$	$0.0001 \\ 0.0001$	$0.0154 \\ 0.0151 \\ 0.0791$	$0.0001 \\ 0.0001$	1.00 0.99 2.17	$\begin{array}{c} 0.00\\ 0.00 \end{array}$	505 715 319 556 113 424	105 111	
Random A Random B				$-0.0001 \\ -0.0003$	$\begin{array}{c} 0.0001 \\ 0.0001 \end{array}$	$0.0155 \\ 0.0155$	$\begin{array}{c} 0.0001 \\ 0.0001 \end{array}$	$\begin{array}{c} 1.00 \\ 0.99 \end{array}$	$\begin{array}{c} 0.00 \\ 0.00 \end{array}$	412 069 127 015	95 68	
Leicestershire												
Ferns (survey) Random A Pandom B	21	535	400	0.1116 0.0002	0.0006	0.1259 0.0433 0.0432	0.0004	3.26 1.01	0.01	1013 3517 1860	4	
Carex (survey) Random A	28	535	400	0.1221 0.0002	0.0005	0.0432 0.1146 0.0435	0.0005	3.64 1.01	0.01	1240 3469	4	
Random B Leguminosae (survey)	42	535	400	$-0.0003 \\ 0.0599$	0.0004	$0.0432 \\ 0.0848$	0.0004	1.00 3.81	0.01	1846 1447	5	
Random A Random B Scrophulariaceae	35	535	400	$0.0006 \\ 0.0002 \\ 0.0466$	$0.0002 \\ 0.0004$	$0.0432 \\ 0.0429 \\ 0.0839$	$0.0003 \\ 0.0005$	1.03 1.00 2.74	$\begin{array}{c} 0.01\\ 0.02 \end{array}$	12 292 2169 2705	5 5	
(survey) Random A Random B				$-0.0004 \\ 0.0004$	$0.0004 \\ 0.0004$	0.0435 0.0432	0.0003 0.0003	0.99 1.02	0.01 0.01	10 198 3329	9 5	
Common herbs (survey)	175	535	400	0.0539		0.0945		10.48		7015		
Random A Random B	25		100	0.0001	$0.0001 \\ 0.0001$	0.0433 0.0433	$0.0001 \\ 0.0000$	1.02 1.01	$\begin{array}{c} 0.01 \\ 0.02 \end{array}$	9543 7942	2 2	
Random A Random B	35	535	400	$ \begin{array}{r} 0.0463 \\ -0.0006 \\ 0.0002 \end{array} $	$0.0005 \\ 0.0004$	$0.0868 \\ 0.0435 \\ 0.0434$	$0.0003 \\ 0.0004$	2.75 0.98 1.01	0.02 0.01	13386 61543 16495	55 24	

TABLE 1. Co-distribution in biological surveys of plant communities, compared with randomized records.

Notes: Columns to the right give the estimates of the binary correlation coefficient r_{species} , its standard deviation $s(r_{\text{species}})$, the V statistic of Pielou and Robson (1972), and the C score of Stone and Roberts (1990). The first row for each plant group is the survey statistic. The other two rows give the mean and standard error for 20 independent randomizations of the survey data, either unconstrained (random A) or constrained so as to maintain overall species range (number of records; random B). "Common herbs" excludes species recorded from <50 or >350 grid squares.

ronment, or strong selection in an uncorrelated heterogeneous environment, do not yield any strongly marked non-neutral pattern. Even strong selection in a correlated environment cannot be reliably distinguished from neutrality, when the landscape is fine grained, with habitats represented by many small patches. It is only strong selection acting in a coarse-grained environment that yields a markedly non-neutral outcome, which is detected most reliably by the very large value of *C*. Any model that incorporates environmental covariance will produce similar results. Thus, the properties of a gradient model are similar to those of the habitat model in a correlated environment, with measures of association increasing as the standard deviation of the fitness function becomes smaller.

Co-distribution depends on the scale of the survey

Neighboring sites can be grouped together so as to form blocks of any size, and patterns of variance and

covariance are liable to change with block size (Pemadesa et al. 1974). In neutral communities, most pairs of species yield a loose lower-triangular graph for abundance at any blocksize (Fig. 3). In uncorrelated heterogeneous environments, species adapted to the same substrate show a strictly L-shaped graph at small block size: Most sites have neither species, whereas those that have one lack the other, regardless of the extent of specialization and the strength of selection (Fig. 4). As block size increases, this breaks down to give loosely lower-triangular graphs resembling those of neutral communities. In correlated environments, the results depend on whether a pair of species are adapted to substrates that occur in the same habitat or in different habitats. If they are adapted to substrates in different habitats, the co-distribution graph is loosely Lshaped, breaking down at large block sizes (Fig. 5). For species adapted to the same habitat, the observations form three clusters of points when unit sites are

No.									
sites	Birth	Death	Dispersal	Immigration	Model	r _{species}	s(r)	V	С
3600	0.51	0.5	0.1	0.001	neutral	-0.00757	0.1275	0.698	634 448
					random A	0.00011	0.0169	1.005	720 580
					random B	0.00003	0.0166	1.002	610 701
2500	0.65	0.5	0.05	0.001	neutral	-0.00935	0.0765	0.530	164 227
					random A	-0.00029	0.0200	0.986	169 759
					random B	0.00017	0.0199	1.008	159 900
2500	0.7	0.5	0.25	0.00001	neutral	-0.00360	0.1093	0.818	338 659
					random A	0.00003	0.0200	1.002	387 491
					random B	-0.00022	0.0202	0.989	347 986
225	0.8	0.5	0.1	0.001	neutral	-0.00405	0.1567	0.757	971
					random A	-0.00028	0.0668	0.990	1932
					random B	0.00080	0.0666	1.052	1529
225	0.8	0.5	0.05	0.0001	neutral	-0.00816	0.1935	0.828	3193
					random A	-0.00011	0.0673	1.002	4471
					random B	0.00025	0.0670	1.022	3034
225	0.8	0.5	0.95	0.01	neutral	0.07401	0.0939	4.668	694
					random A	-0.00010	0.0666	1.000	3409
					random B	-0.00007	0.0667	0.999	938
225	0.3	0.1	0.1	0.01	neutral	-0.00266	0.1165	0.742	1842
					random A	0.00003	0.0662	1.005	2177
					random B	0.00005	0.0666	1.000	1791
225	0.8	0.6	0.01	0.1	neutral	0.18570	0.1181	9.448	938
					random A	0.00016	0.0665	1.013	1433
					random B	0.00012	0.0663	1.010	1344

TABLE 2. Co-distribution in neutral community models (NCM), compared with randomized records.

Notes: Each of the eight cases refers to a single realization of a neutral community model with a pool of 50 species and a maximum density of 100 individuals/site, run for 2000 cycles (10 000 cycles for the first case). Measures of association (see Table 1) were calculated for the neutral model output and for 20 independent randomizations of this output, either unconstrained (random A) or constrained so as to maintain overall species range (number of records; random B). Standard errors of the measures are similar to those given in Table 1.

sampled (Fig. 6A). Most sites have neither species. Others are dominated by the species adapted to their substrate, but often have substantial numbers of the other species. These are recent immigrants, which are more frequent than expected because of the proximity of sites bearing substrates in the same habitat. If sites are blocked in pairs, the observations now form six clusters of points (Fig. 6B). This arises as follows. Let the two substrates belonging to the same habitat, to which the two species are adapted, be A and B, and

A) Parameters												
Modified value	If reduced				If increased					Range		
K'			xК		9 <i>xK</i>					0-9K		
b'			xb		b + x(1 - b)					0-1		
d'			xd		d + x(1 - d)					0-1		
μ'		exp(1	n m - x)	$\exp(\ln m + x)$					0.37–2.7 m		
m'			хи		u + x(1 - u)					0-1		
B) Results												
		r			s(r)			V			С	
Model	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
Random	0.0001	0.0020	0.0003	0.0669	0.0014	0.0002	1.01	0.10	0.01	305	4	1
Neutral	-0.0017	0.0091	0.0013	0.0884	0.0405	0.0057	0.82	0.51	0.07	537	557	79
Weak selection	-0.0033	0.0031	0.0004	0.0872	0.0260	0.0037	0.70	0.23	0.03	342	219	31
Strong selection												
Uncorrelated	-0.0045	0.0091	0.0013	0.0955	0.0373	0.0053	0.73	0.52	0.07	430	421	60
Fine-grain correlated	-0.0058	0.0041	0.0006	0.0910	0.0275	0.0039	0.63	0.34	0.05	231	174	25
Coarse-grain correlated	-0.0028	0.0136	0.0034	0.1192	0.0429	0.0107	0.81	0.68	0.17	61 337	61 044	15 261

TABLE 3. Effect of local selection on co-distribution.

Notes: Each model had the following kernel: number of species (N) = 50; number of sites (M) = 225; limiting density (K) = 50 individuals/site; birth rate (b) = 0.5; death rate (d) = 0.3; local dispersal (u) = 0.1; immigration (m) = 0.001 per species. Fifty realizations of each model were run. In each realization, the parameters K, b, d, u, and m were each independently increased or decreased with equal probability. Each was then assigned a modified value obtained from uniform random numbers x, independently for each parameter in each realization of each model. There is a constraint such that if b' < d', then these values are exchanged.



FIG. 3. Bivariate scattergrams from a neutral community model at four levels of survey grain.



FIG. 4. Bivariate scattergrams from an uncorrelated environment model with strong selection. The parameter set is that for the first model of Table 2, with s (selection coefficient) = 0.25. Species numbers are arbitrary codes.



FIG. 5. Bivariate scattergrams from a correlated environment model, for species adapted to substrates belonging to different habitats. The parameter set is that for the first model of Table 2, with h = 3 substrates per habitat, H = 15 habitats, and s (selection coefficient) = 0.25. Species numbers are arbitrary codes. The 3600 sites were aggregated into contguous square blocks with a block size of 1, 4, 25, or 100.

let X designate a substrate belonging to any other habitat. The six clusters of points correspond to pairs of sites with the combinations XX, AX, BX, AA, BB, and AB. Larger blocks, forming patches smaller than a habitat, give more complex clusters of points on the same principle (Fig. 6C), although in practice, sampling error soon smears the points over the lower triangle of the diagram, to which they are confined because of the fixed upper limit on the total number of individuals within a patch (Fig. 6D, E). When the block corresponds with the habitat, the graph separates into two groups of points, one of which represents patches where neither species is present (except as rare immigrants) and the other patches where both species are present within the habitat that includes both the substrates to which they are adapted (Fig. 6F). Among the patches where both species occur, their abundance is negatively, not positively, correlated: This is because of the fixed capacity of the patch. (Note that the graph for species occupying different habitats becomes strictly L-shaped at this scale, because neither occurs [except as rare migrants] in blocks occupied by the other.) For blocks of sites larger than the habitat, these two groups of points fuse, because the blocks of sites now span several different habitats (Fig. 6G). With increasing block size, this cloud of points becomes more and more linear (Fig. 6H), because the blocks vary stochastically

in the number of patches of given habitat they contain; those that happen by chance to contain few patches of a given habitat will have few individuals of those species adapted to its constituent substrates, whereas those that have more are likely to have many individuals of any two well-adapted species. To summarize these results: When abundance has been measured, strong selection in coarse-grained environments creates strong positive correlations between species adapted to the same habitat, although this will be detected only in over-blocked surveys. In all other cases, survey results cannot be distinguished from neutral models because the relationship between grain size and patch size is unknown.

Prior classification of species and sites filters out a functional signal

Because all individuals regardless of species membership are equivalent at all sites in a neutral model, all prior classifications of species and sites are likewise equivalent. In practice, we can group species into taxa on genetic criteria, and sites into habitats on ecological criteria, in advance of conducting a survey. The effect of these classifications should be to redistribute the covariance so that it become larger for comparisons within categories. Thus, $r_{\rm species}$ should be greater for related species belonging to the same taxon, and $r_{\rm sites}$



FIG. 6. Bivariate scattergrams from a fine-grained correlated environment model, for species adapted to substrates belonging to the same habitat. The parameter set is that for the first model of Table 2, with h = 3 substrates per habitat, H = 15 habitats, and s (selection coefficient) = 0.25.

should be greater for similar sites belonging to the same habitat.

The Leicestershire grid survey has low positive covariance over all pairwise combinations of species. Con When plant families are analyzed separately, this value is consistently larger. Furthermore, when comparisons within a family are restricted to congeneric species, it is larger still; roughly speaking, the average value of r_{site}

 $r_{\rm species}$ increases from ~0.02 for unrelated species to ~0.2 for pairs of species in the same genus (Table 4).

The Leicestershire habitat survey has a much smaller correlation among sites than the grid survey. When two contrasting habitats are chosen, such as grassland and woodland, comparisons between sites belonging to the same habitat show a marked increase in the value of r_{sites} , whereas comparisons between sites representing

		All com	parisons	No. species	Compa betweer	arisons 1 genera	Compa within	arisons genus
Plant group	No. species	Mean	SE	pairs	Mean	SE	Mean	SE
All species	993	0.0231	0.0001					
Caryophyllaceae	31	0.0717	0.0044	5	0.0791	0.0134	0.2114	0.0385
Cruciferae	44	0.0790	0.0027	6	0.0801	0.0085	0.1664	0.0702
Leguminosae	45	0.0519	0.0026	5	0.0484	0.0109	0.1180	0.0404
Compositae	94	0.0717	0.0012	9	0.0953	0.0170	0.2630	0.0443
Gramineae	73	0.0377	0.0015	10	0.0536	0.0052	0.1090	0.0518

TABLE 4. Co-distribution under a taxonomic classification of species.

Notes: Values are binary correlations between pairs of species calculated for the Leicestershire grid survey. The first set of columns shows the results for all pairwise combinations of species. Congeneric species pairs were then selected from each family, using those pairs with intermediate range (number of records) when more than two were available. The second set of columns shows the results for these species pairs and for all pairwise combinations of species from different genera within a given family.

different habitats show a corresponding decrease (Table 5). Roughly speaking, the effect of ecological classification is, in this case, to increase r_{sites} from ~0.03 for sites belonging to different habitats to ~0.3 for sites belonging to the same habitat.

DISCUSSION

Species associations and assembly rules

Most of the quantitative analysis of species co-occurrence in the ecological literature of the last 25 years has stemmed from the influential article on "assembly rules" by Diamond (1975). He found consistent patterns in the composition of bird communities on islands, especially the absence of some combinations of species that were expected to occur (that is, an excess of negative covariance of species over sites). This was interpreted as evidence that one species may exclude another through competition because of their ecological similarity. The obverse is that species can coexist if they are sufficiently dissimilar, through a process of local adaptation to heterogeneous landscapes, leading to a functional interpretation of community composition. The analysis was challenged by Connor and Simberloff (1979), who showed that randomly assembled communities displayed similar patterns, which could not, therefore, be cited as evidence for the occurrence of competition. This set off a long controversy (see Weiher and Keddy 1999), in which surveys, often of rather dubious quality, were analyzed by a rich variety of statistical procedures in an attempt to confirm or deny the existence of patterns of species co-occurrence that would testify to the effect of competition in structuring communities. A recent analysis of nearly one hundred surveys concluded that communities of birds, mammals, ants, and plants showed clearly nonrandom degrees of co-occurrence, whereas fishes, herpetofauna, and most invertebrates did not (Gotelli and McCabe 2002). The nonrandom patterns were consistent with Diamond's (1975) original formulation of assembly rules, and thus with the importance of competition in determining the composition of communities, while showing that these conclusions failed to apply to some groups of organisms for physiological or ecological reasons yet to be determined (see also Gotelli and Rohde 2002). It is now clear, however, that highly nonrandom patterns of co-distribution can arise in communities of species with identical properties.

Landscape structure and co-distribution

Local selection in a heterogeneous environment generates correlations between species to an extent that depends on the intensity of selection, up to the limit where species are almost completely segregated into perfectly distinct habitats because each is lethal in all habitats save one. Suppose that there are M sites comprising H equally frequent habitats: All sites belonging

TABLE 5. Co-distribution under an ecological classification of sites.

		Corre	Correlation between sites				
Comparison	No. sites	Mean r _{sites}	$s(r_{\rm sites})$	$SE(r_{sites})$			
All grid squares	535	0.5646	0.0588	0.0002			
All habitat sites	103	0.1648	0.1455	0.002			
Grassland + forest sites	40	0.1834	0.1958	0.007			
Grassland sites only	19	0.4204	0.1119	0.0086			
Forest sites only	21	0.2804	0.1442	0.01			
Grassland vs. forest	40	0.0308	0.0798	0.004			

Notes: Values are binary correlations between pairs of sites calculated for the Leicestershire survey. "Habitat sites" are separately surveyed sites, typically a few hectares in extent, identified as representative of a particular habitat. Twelve habitats were recognized in all, of which grassland and forest were the most numerous.

to the same habitat have the same combination of species, and no species occurring in a given habitat occurs in any other habitat. There are H(N/H)(N/H - 1)/2 comparisons between species adapted to the same habitat, for which $r_{\text{species}} = +1$, whereas the remaining comparisons are between species adapted to different habitats, which have $r_{\text{species}} = -(M/H)/[M - (M/H)]$. The mean and parametric variance of the site associations are then

$$Mean(r_{species}) = -1/(N - 1)$$

$$Var(r_{species}) = \frac{N^{2}(N - H)}{(N - 1)^{3}H} + \frac{N(H - 1)[(H - 1) - (N - 1)]^{2}}{H(N - 1)(H - 1)^{2}(N - 1)^{2}}$$

The variance is zero for N = H (a single species adapted to each habitat), and is approximately 1/(H - 1) for $N \gg H > 1$. A similar result holds for the site associations, replacing the number of species N by the number of sites M, and the number of habitats H by the number of sites in which each species is found M/H.

The degree to which these extremes are realized, however, depends on the configuration of substrates within habitats. The simplest case is one in which habitats consist of a single substrate, and a single species is well adapted to each site. Each site will then be occupied by its diagnostic species, together with recent immigrants from neighboring sites. If substrates are allocated to sites at random, then the pool of available immigrants for any given target site will be a random sample of the overall species pool. Consequently, no consistently extreme species associations, positive or negative, are expected to emerge. If nearby sites are more likely to have the same substrate, then recent immigrants are more likely to be the same species as the residents. This will resemble the effect of stronger local selection, but it will not create any more pronounced species associations. If several species are well adapted to any given substrate, this may create species associations, if species adapted to the same substrate are consistently found together, whereas species adapted to different substrates are consistently found apart. This is not necessarily the case, however: In these circumstances, site and species associations may be indistinguishable from those generated by a neutral model, regardless of the strength of local selection or the rate of local dispersal. The reason for this surprising result is that the set of species well adapted to any given substrate has neutral dynamics; if site capacity is small and local dispersal low, each site bearing a given substrate will almost always be occupied by only one of the species adapted to that substrate. This explains the L-shaped abundance plots for similarly adapted species shown in Fig. 4.

Consistently extreme species associations thus require more than environmental heterogeneity and local selection. They require, rather, that members of a specific set of differently adapted species are usually found in nearby sites. This will happen if substrates are spatially correlated, so that the environment consists of an array of patches, each being made up of sites representing one of a relatively small number of combinations of substrates. Each combination of h substrates constitutes a habitat. Because certain substrates are usually found together in the same patch, the corresponding combination of species will also tend to occur together. Co-occurrence thus requires environmental covariance rather than merely environmental variance.

The strength of species associations in a correlated environment depends on how many species are specialized to a given substrate. When a single specialist has evolved for each substrate, the associations are strong because each habitat supports the same combination of specialists. When there are several specialists available for each substrate, these associations are weakened, because the neutral dynamics of equally well-adapted species will create many possible combinations for each habitat, only one of which will be realized in a given patch. Consequently, species associations in correlated environments with substitutable specialists may be scarcely distinguishable from neutral communities.

Survey design

The way in which the original records of a survey are aggregated into blocks has a very consistent effect on species associations: Regardless of environmental structure or the strength of selection, the variance of species associations generally increases with block size. The reason is that any given site is likely to contain only a small fraction of the species in the regional pool, provided that the landscape and the survey are reasonably fine grained and local dispersal is not very high. At the scale of the site, therefore, most species are likely to be rare. As block size increases, the combined species list grows, and an increasing number of species are recorded at intermediate range, thereby increasing the variance of species associations.

If the grain of the survey (block size) is large relative to the true grain of the environment (sites) most of the environmental variation found over the extent of the survey would then be found within the grain of the survey, and few strong associations could be detected because most blocks would include sites representing all available habitats. Roughly speaking, this requires that landscapes are very fine grained, to the extent that most of the environmental variation found in a unit area is also found in 1/1000 of that area. It is certainly true that one might find woodland, heath, and pasture within 4 km² of an English county; or seasonal runoff channels, tip-up mounds, and bare rock surfaces within 1 ha of a North American old-growth forest. Whether or not this is quantitatively plausible does not seem to have been ascertained in general, although it is not

implausible that the combinations of substrates that constitute habitats should occur at scales below that of most biological surveys. Edaphic variation at spatial scales of 1-50 m, for example, has been documented extensively at Mont St.-Hilaire (Bell and Lechowicz 1991, Lechowicz and Bell 1991). Nevertheless, the objection does not seem very powerful. In the first place, if it were true, one would expect to find almost all species in almost all survey blocks, which is not the case. Secondly, the simulations show that strong positive correlations are most likely to be observed in overblocked surveys. Finally, the objection amounts to admitting that any signal contributed by local adaptation may readily be obscured by blocking, through the unknown relation between the grain of the survey and the grain of the environment. Nevertheless, it is worthwhile to ask whether very fine-scale surveys do, as a matter of fact, show anomalously strong associations.

Very fine-grained surveys are exceedingly laborious and are therefore rarely attempted. One exception is a survey of precisely located trees before settlement in the southern boreal forest of Minnesota, analyzed by Friedman et al. (2001). The data concerned 9280 small plots each containing four trees, distributed over an area of 3.2×10^6 ha, effectively combining small grain with large extent. There are 36 possible pairwise combinations of the nine abundant tree species, of which 32 were observed. Species co-occurrence was expressed by the Jaccard index $J = n_{11}/(n_{11} + n_{10} + n_{01})$. The value of this index was never high, having a maximum of 0.24 and being < 0.1 in 22 out of 32 cases. Nine rare species showed uniformly low degrees of cooccurrence. On the other hand, most plots contained three or four individuals of the same species. These data are certainly not random, but the generally low level of co-occurrence and the aggregation of species are consistent with the output of spatial NCMs. For example, the Jaccard coefficient for the fine-grained neutral community model at the head of Table 2 has a mean value range of 0.11-0.13 (SD range of 0.08-0.1) in replicate runs.

Range and co-distribution

Whether or not local selection is acting, the strength of species associations will depend on the distribution of species range. This is because common species that are found everywhere, or rare species that occur at very few sites, cannot show high levels of association, positive or negative, with other species. Under a simple model of local selection in an heterogeneous environment, the range of a species will depend on the number of sites to which it is well adapted. So far, it has been assumed that this number is drawn from a uniform distribution, so there is little variation in range among species, and most will occur in many sites if the number of habitats H is not too large, selection s not very strong, and dispersal u moderately large. If it is drawn instead from a geometric distribution, we can create a

landscape in which habitats differ widely in frequency, causing a corresponding divergence in species ranges. This reduces the strength of species associations to very low values. Species associations may be strongly influenced by this effect. The surveys that I used have coefficients of variation of range in the region of 2, reflecting large differences between very widespread species found almost everywhere and very rare species restricted to one or a very few sites. Nevertheless, restricting the analysis to species of intermediate range did not uncover a marked increase in the strength of associations. Moreover, adjusting rates of birth and dispersal can generate wide variation of species range in neutral communities, leading to a similar reduction in the strength of associations, so there is no distinctive difference between neutral and locally adapted communities in their response to variation in species range.

The limits of theory

Despite the difficulty of demonstrating any distinctive difference between survey data and neutral community models, it is nevertheless very unlikely that plant communities such as those analyzed here are truly neutral. Classifying species into taxa, or sites into habitats, produces a consistent increase in covariance that confirms everyday field experience and is inconsistent with neutrality. The main conclusion of the analysis is not that local selection and local adaptation do not occur, but rather that they cannot usually be unambiguously detected from unclassified survey data, however extensive or detailed. Thus, the clustering of functionally similar species is clearly apparent only under intense selection in a coarse-grained environment where environmental factors are strongly correlated. In other circumstances, the systematic effect of local selection may be over-ridden by local dispersal and stochastic drift.

This interpretation implies that the output of neutral community models is a good first approximation to survey data because the models themselves are good first approximations to community processes. There are bulk properties of communities that can be evaluated without regard to the identity of the species concerned. These include, for example, the distribution of abundance, the overall species diversity, and the average strength of species co-distribution. The degree to which these bulk properties depart from randomness is often adequately predicted by neutral models because it is indeed largely attributable to drift and dispersal. On the other hand, there are particular attributes of communities that depend on species identity. These would include, for example, the difference in species composition between habitats, or the difference in habitat preference between species. Provided that these attributes are based on prior categories, they must be interpreted in terms of local selection among divergently specialized species. A similar dichotomy exists in other sciences. The nature of interactions between molecules in a gas will depend on the chemical nature of the molecules involved, whereas the kinetic law of gases predicts bulk properties without requiring a knowledge of chemical composition. Somewhat closer to home, describing the mass balance of phosphorus in an estuarine ecosystem would normally require no more than the crudest categorization of the organisms involved, paying little if any heed to variation in the phosphorus economy among the innumerable species in the community. In some cases, this heuristic principle of equivalence will break down: The presence of zebra mussels rather than unionid clams, for example, may make a very large difference to the dynamics of phosphorus in a lake. In most places and at most times, however, the bulk properties of communities may be insensitive to the details of interactions among species, and are then adequately predicted, and correctly explained, in terms of neutral processes.

The converse conclusion is that the bulk properties of communities are not likely to convey much information about the details of species interactions. Randomization is clearly an inappropriate null hypothesis for testing functional interpretations of co-distribution, because neutral models yield strikingly nonrandom outcomes. It is instead the neutral model itself that should be used as a null hypothesis; but it seems that it will seldom be possible to reject it, unless in exceptional circumstances, even though local selection is widespread. From this point of view, unclassified survey data, although useful for describing communities, is of little use for analyzing them (see Lawton 1999).

Acknowledgments

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and by Le Fonds québécois de la recherche sur la nature et les technologies.

LITERATURE CITED

- Bell, G. 2000. The distribution of abundance in neutral communities. American Naturalist 155:606–617.
- Bell, G. 2001. Neutral macroecology. Science **293**:2413–2418.
- Bell, G. 2003. The interpretation of biological surveys. Proceedings of the Royal Society of London, Series B 270: 2531–2542.
- Bell, G., and M. J. Lechowicz. 1991. The ecology and genetics of fitness in forest plants. I. Environmental heterogeneity measured by explant trials. Journal of Ecology 79: 663–685.
- Bell, G., M. J. Lechowicz, and M. J. Waterway. 2001. The precision of adaptation in forest plants. Pages 117–138 *in* J. Silvertown and J. Antonovics, editors. Integrating ecology and evolution in a spatial context. Blackwell Science, Oxford, UK.
- Braun-Blanquet, J. 1932. Plant sociology: the study of plant communities. English translation of Pflanzensoziologie by G. D. Fuller and H. S. Conard. McGraw-Hill, New York, New York, USA.
- Chave, J. 2004. Neutral theory and community ecology. Ecology Letters 7:241–253.
- Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology 60: 1132–1140.

- Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.
- Freedman, D. A. 1983. A note on screening regression equations. American Statistician 37:155–158.
- Friedman, S. K., P. B. Reich, and L. E. Frelich. 2001. Multiple-scale composition and spatial distribution patterns of the north-eastern Minnesota presettlement forest. Journal of Ecology 89:538–554.
- Gilbert, B., and M. J. Lechowicz. 2004. Neutrality, niches and dispersal in a temperate forest understory. Proceedings of the National Academy of Sciences (USA) 101:7651– 7656.
- Gilpin, M. E., and J. M. Diamond. 1982. Factors contributing to non-randomness in species co-occurrences on islands. Oecologia 52:75–84.
- Goodall, D. W. 1953. Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. Australian Journal of Botany 1:39–63.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621.
- Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology 83:2091–2096.
- Gotelli, N. J., and K. Rohde. 2002. Co-occurrence of ectoparasites of marine fish: a null model analysis. Ecology Letters 5:86–94.
- Greig-Smith, P. 1961. Data on pattern within plant communities. I. The analysis of pattern. Journal of Ecology 49: 695–702.
- Hubbell, S. P. 1995. Toward a theory of biodiversity and biogeography on continuous landscapes. Pages 171–199 *in* G. R. Carmichael, G. E. Folk, and J. E. Schnoor. Preparing for global change: a mid-western perspective. Academic Publishing, Amsterdam, The Netherlands.
- Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.
- Kershaw, K. A. 1960. The detection of pattern and association. Journal of Ecology 48:233–242.
- Lawton, J. H. 1999. Are there general laws in ecology? Oikos **84**:177–192.
- Lechowicz, M. J., and G. Bell. 1991. The ecology and genetics of fitness in forest plants. II. Microspatial heterogeneity of the edaphic environment. Journal of Ecology 79: 687–696.
- MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA.
- Pemadasa, M. A., P. Greig-Smith, and P. H. Lovell. 1974. A quantitative description of the distribution of annuals in the dune system at Aberffraw, Anglesey. Journal of Ecology 62:379–402.
- Pielou, E. C., and D. S. Robson. 1972. 2 to k power contingency tables in ecology. Journal of Theoretical Biology 34: 337–352.
- Primavesi, A. L., and P. A. Evans, editors. 1988. Flora of Leicestershire. Publication Number 89, Leicestershire Museums, Leicester, UK.
- Schluter, D. 1984. A variance test for detecting species associations, with some example applications. Ecology 65: 998–1005.
- Simberloff, D., and E. F. Connor. 1979. Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. Pages 123–138 in G. P. Patil and M. L. Rosenzweig, editors. Contemporary quantitative ecology and related ecometrics. International Cooperative Publishing House, Fairland, Maryland, USA.

- Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85:74-79.
- Weiher, E., and P. Keddy. 1999. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press, Cambridge, UK. Wiens, J. A. 1981. Single-sample surveys of communities:
- are the revealed patterns real? American Naturalist 117: 90-98.
- Williams, W. T., and J. M. Lambert. 1959. Multivariate methods in plant ecology. I. Association-analysis in plant communities. Journal of Ecology 47:83-101.
- Williams, W. T., and J. M. Lambert. 1961. Multivariate methods in plant ecology. III. Inverse association-analysis. Journal of Ecology **49**:717–729. Wilson, J. B. 1991. Does vegetation science exist? Journal
- of Vegetation Science 2:289-290.