

Individual-centered analysis of mapped point patterns representing multi-species assemblages

Podani, János & Czárán, Tamás

Department of Plant Taxonomy and Ecology, L. Eötvös University, Ludovika tér 2, H-1083 Budapest, Hungary;
Fax +36 1 3338764; E-mail podani@ludens.elte.hu and czaran@ludens.elte.hu

Abstract. On the basis of Ripley's combined count-distance method, Juhász-Nagy's information theoretical functions and the proposition of Williams et al. for the study of small-scale community pattern, a new procedure is suggested for elucidating multi-species point patterns based on digitized field data. The method utilizes nested circular plots with increasing radii drawn around each individual and determines changes in floristic composition along this space series. The information provided by detecting the species composition around the sample plant is calculated, and its observed mean for all individuals is compared, for each radius, to the expectation under the null model, i.e. for complete spatial randomness of all points. The departure from randomness is illustrated by conventional profile diagrams and is tested for significance based on confidence envelopes simulated by Monte Carlo methods. One advantage of the individual-centered sampling strategy is that the role of each species in influencing its own neighbourhood can be analyzed separately, providing information for the assessment of guild structure and assembly rules in communities. The performance of the method is evaluated using artificial and simulated point patterns.

Keywords: Compositional diversity; Information theory; Null model; Simulated community; Spatial analysis.

Introduction

Currently, one of the most popular and mathematically best understood techniques used for the analysis of species pattern is that of Ripley (1981), known as *K*-function or second-order analysis of point patterns. The sampling technique involves the placement of a circular sample plot of radius *t* around individuals and counting the number of neighbouring individuals within the circle. Then, deviation of the observed number from expectation (referring to the null situation of complete spatial randomness, CSR; Diggle 1983) is computed. The calculation is repeated for a range of *t*-values to examine scale dependence of the statistic, and to depict a particular area at which the point pattern exhibits maximum deviation from CSR towards either aggregation or segregation. Applications of Ripley's method to

ecological problems are many; see Haase (1995) for a recent account on the topic.

The procedure has been extended to the two-species case (Harkness & Isham 1983). In this bivariate second-order spatial analysis the number of individuals of species 2 found within distance *t* of an individual of species 1 is determined, and *vice versa*. The deviation of the observed mean from random expectation (under the assumption of CSR) is examined for various values of *t*. The spatial scale at which the two species are maximally associated ('attraction') or dissociated ('repulsion') can thus be identified. If the deviations remain within the confidence envelope generated for random counterparts of the same point pattern, the two species may be considered spatially independent for the given range of *t*-values. Bivariate spatial analysis has received several ecological applications (e.g. Cox 1987; Kenkel 1994; Rebertus et al. 1989).

The question arises how we can extend this procedure to several species. Can we identify the spatial scale at which all or most species show maximum deviation from random expectation? We will show that the information-theory approach by Juhász-Nagy (1976, 1993; Juhász-Nagy & Podani 1983) can easily be brought into a format comparable with the above-mentioned univariate and bivariate techniques. Such multivariate point pattern analysis can be linked with the debate on 'community assembly rules' and their interpretation in terms of 'niche limitation' (e.g. Wilson 1994; Zobel et al. 1993; Bartha et al. 1995). It is clear that a theoretically sound and well-understood methodology, which is demonstrated on a range of actual and simulated data, is required before coming to any conclusion regarding the ecological significance and interpretation of any spatial statistic.

Preliminary considerations

A simple multivariate approach would involve computation of all possible pairwise (bivariate) deviates and their incorporation into a single, 'community level'

statistic (e.g. a mean) for each value of t . Such an approach would meet at least three serious difficulties: (1) There is a strong interdependence of pairwise measures: the statistic for species 1 and 2 and that for species 2 and 3 are not independent and they determine in some way the statistic for species 1 and 3. (2) Species with high density would dominate the analysis at the expense of the uncommon species which may convey more ecological information. (3) The interpretation of overall repulsion and association of species is impossible because the pairwise measures can easily cancel out each other. Note that Ripley (1987, pp. 424-426) is quite sceptical about extensions to the multi-species situation and attributes even limited success to the bivariate analysis.

It would make more sense if a single overall measure of the deviation from randomness could be found which is less sensitive to rarity vs. commonness. To achieve this, we have to compromise by giving less emphasis to the number of individuals so important in the uni and bivariate cases, and *partly* reduce the problem to the presence/absence situation. Although this may seem to have no direct relationship with Ripley's ideas, many ecologists would agree that with larger numbers of species, a presence/absence approach will better reveal environmental heterogeneity than absolute quantities at any scale. This has been implicitly assumed by attempts to explain niche limitation by the variance of the number of species (e.g. Schluter 1984; Palmer 1987; Zobel et al. 1993; Wilson & Watkins 1994). Ripley (1987) also makes this shift to the presence/absence case when discussing analyses of multi-species point patterns. However, species densities will not be overlooked in our approach: they will be used to compute the random expectation to which actual point patterns will be compared.

Historical background

Compositional diversity

A review of analyses of multivariate plant pattern (Podani et al. 1993) revealed that the most sophisticated approach to date is that of Juhász-Nagy (1976). He applied a family of information theoretical functions to a series of data matrices, each summarizing presence/absence scores of species in a set of sample plots of a given size. The most interesting function is Shannon's entropy measure, used in a non-conventional way. Contrary to the usual diversity computations based on the empirical frequency distribution of species, Shannon's entropy is applied to the frequency distribution of species combinations (florulas) thus measuring the so-called *compositional* (or *florula*) *diversity* (Juhász-Nagy & Podani 1983; Juhász-Nagy 1993) of the community.

This diversity measure is strongly scale-dependent and shows a peaked effect when displayed against plot size. The area pertaining to its maximum value reflects the scale point at which the community reaches its highest variation, thus considered optimal for community characterization. Changes of this maximum area over time have been used, for example, to evaluate different aspects of secondary succession in abandoned fields (Juhász-Nagy & Podani 1983) and of revegetation processes on dumps of open strip mines (Bartha 1992).

Pooled entropy

In a sampling experiment in rain forests Williams et al. (1969) used the '*information content*' (e.g. Sneath & Sokal 1973), a measure included in Juhász-Nagy's family of functions. This function is the pooled entropy of the species, i.e. the arithmetic sum of Shannon's entropies calculated separately for each species based on their presence and absence in sampling units. Williams et al. (1969) used a sampling technique based on point clumps: each plant, together with its *v* '*multiple nearest neighbours*' were considered as a sampling unit. The pooled entropy was calculated for samples based on increasing values of v . The authors found that the pooled entropy is in fact the function of the mean distance from the reference individuals, i.e. it is distance and, consequently, area that matter rather than v . Then, Podani et al. (1993) developed a pattern analysis technique, in which each individual is considered as the centroid of a circular plot with radius t , and the florula of these plots is determined for increasing values of the radius. This is the sampling procedure advocated in the present paper.

Species associations

Williams et al. (1969) evaluated the changes of pooled entropy over increases of v in multiple nearest neighbour analysis and demonstrated its peaked effect, but they did not place the results into an appropriate context. It was Juhász-Nagy who recognized that pooled entropy – '*local distinctiveness*' in his terminology – equals compositional diversity when there are no associations among the species, because the following simple relationship (for mathematical formalism, see Juhász-Nagy & Podani 1983; Bartha 1992) holds:

$$\text{compositional diversity} = \text{pooled entropy} - \text{overall species associations}$$

As mentioned above, the first two quantities have maxima along the spatial scale, and hence *overall species associations* (or '*associatum*', sensu Juhász-Nagy 1976) will also have at least one maximum. With a sampling procedure involving a series of increasing

plots we will therefore be able to recognize the scale point at which maximum overall (negative and positive) associations exist. Maxima for the three functions are found within a relatively short range along the spatial scale, and the size of this range as well as the sequence of the maxima ('characteristic ordering') within that range are typical (Juhász-Nagy & Podani 1983 gave examples). In actual communities, there are always associations among species, and in randomly simulated species assemblages some associations, even 'significant' ones will inevitably emerge (Schluter 1984). Therefore, the associatum must be treated with caution, and hence the subsequent proposition in this paper is to base pattern analysis solely on a quantity closely related to compositional diversity and its expectation.

Variance in species number

Mathematically, compositional diversity is the joint entropy for a 2^s contingency table, where s is the number of species in the studied region. The analysis of such contingency tables was also mentioned by Pielou (1972), but this had no big impact on subsequent pattern studies because scale dependence was neglected. Since Juhász-Nagy's work was overlooked, the prevailing view was that the 2^s contingency tables are 'impractical' and 'intractable' (e.g. Schluter 1984), and the ratio between observed and expected variance of species number (the latter devised from the binomial distribution) became an index of species association (cf. Robson 1972). Although Schluter (1984) warned that "there is no necessary correspondence between the result of the variance test and any ecological process", subsequent interpretations of the variance in species number did not bother us much, in view of the recent discussion on niche limitation. It is often forgotten that variance in species number is subject to scale factors as well (e.g. Palmer 1987), and should be tested by confidence intervals (Tóthmérész 1994).

A novel procedure of pattern analysis

For the elucidation of small-scale community pattern we describe a new method, consisting of eight steps.

1. Mapping. A rectangular area is delineated in the study site. The exact position and identity of plants are recorded in terms of rectangular coordinates. Clearly, the method is limited to situations where the actual extension of plants can be considered pointwise without much distortion. For procedures of creating such point maps in the field, see e.g. Rohlf & Archie (1978).

2. Computerized sampling. The field data are processed (Podani 1987) and analyzed with the FORTRAN program DARIUS (after the rich Persian king; also a permutation of 'radius'). The program examines all plants and determines the surrounding species composition for each within the specified radius. Plants around which no full circles can be drawn because of overlaps with the boundary line are omitted.

3. Probabilities under the assumption of CSR. At this point it becomes inevitable to introduce abbreviations used in the text. Assume that there are s species in the rectangular area of size A . Species i is represented by n_i individuals, so its density is n_i/A . Let t denote the radius of circles to be drawn around each individual. Then, if the spatial distribution of species i is random, the zero term of the Poisson distribution will give us the probability that a circular plot will not contain that species, i.e.

$$p_{i(t)} = e^{-\lambda} \quad (1)$$

where $\lambda = n_i \pi t^2 / A$, that is, the the mean number of individuals found in an area of πt^2 . Clearly, the probability of presence, i.e. the event that at least one individual of the species appears in the plot, is $1 - p_{i(t)}$.

4. Mean compositional information. For the multi-species case the spatial arrangement of all species has to be examined. This arrangement can be most simply expressed in terms of the species combinations that appear in the sample plots. Species combinations can be described in terms of profiles or binary vectors, denoted by \mathbf{F}_j , of length s in which '0' in position i reflects absence, whereas '1' reflects presence of species i . j runs from 1 to 2^s , the possible number of combinations for s species. Given the assumption of CSR, detection of species combination \mathbf{F}_j around an individual has a probability of

$$P[\mathbf{F}_j]_{(t)} = \prod_{i=1}^s q_{i(t)} \quad (2)$$

where $q_i(t) = p_i(t)$ if species i is absent, and $q_i(t) = 1 - p_i(t)$ if species i is present (keep in mind the big difference between two probabilities, p and P : the first refers to species absences whereas the second reflects probabilities of finding species combinations!). For example, for $s = 4$, $\mathbf{F}_1 = [0, 1, 1, 0]$ and $\mathbf{p}_{(t)} = [.2, .4, .9, .9]$ the probability $P[\mathbf{F}_1]_{(t)}$ is $0.2 \times 0.6 \times 0.1 \times 0.9 = 0.0108$. Information theory defines the information gained by learning the outcome of a particular experiment as the negative logarithm of the probability of the event realized, that is, a measurement of the 'surprise' that combination j was detected is

$$I[\mathbf{F}_j]_{(t)} = -\ln P[\mathbf{F}_j]_{(t)} \quad (3)$$

If \mathbf{F}_1 is manifested, then $I[\mathbf{F}_1]_{(t)} = 4.53$. For a very likely species combination, i.e. $\mathbf{F}_2 = [1, 1, 0, 0]$, however, this information is much less, being $I[\mathbf{F}_2]_{(t)} = 0.94$, whereas for the most unlikely combination $\mathbf{F}_3 = [0, 0, 1, 1]$, the information is the highest, $I[\mathbf{F}_3]_{(t)} = 7.13$.

This example illustrates the essence of our approach. When we repeatedly obtain low information values, i.e. we detect species combinations that are the most common ones under CSR, the species assemblage tends to be close to random. Higher I -values result when less likely species combinations prevail and some species co-occur whereas others are segregated more often than expected.

To conclude, the mean information for all individuals, called the *mean compositional information*, is defined as

$$MCI_{(t)} = \sum_{k=1}^{N'(t)} I[\mathbf{F}_k]_{(t)} / N'(t) \quad (4)$$

where \mathbf{F}_k now refers to the species combination actually found around the k th individual and $N'(t)$ is the number of individuals around which circles of radius t can be drawn without overlapping the boundary. This measure will serve as a characteristic of the compositional pattern of the assemblage. However, there were two deliberate simplifications in the above train of thought:

(a) The individual representing the centre of each sample circle should in fact be omitted from the calculations of probabilities, i.e. the abundance of the given species should be $n_i - 1$. Its effect on the results is obviously negligible for high abundances. Species represented by a single individual must be omitted from the analysis, however. A solution of this problem is to compute the information for each circle by omitting the centre species from the computations, i.e. examining the 'subcombination' for all the other $s - 1$ species within the circle. An argument in favour of this is that the effect of the pattern of the centre species is removed from the analysis, which is not so if the centre species is looked for again in the circle. This subcombination approach therefore seems more attractive theoretically, although the equations above and below become more complicated. Separate $MCI_{(t)}$ values need to be computed for each species and then their overall average computed to obtain $MCI_{(t)\text{sub}}$, (for details, see App. 1). Also, it exceedingly increases computing effort, which may be critical in determining the confidence envelopes described below. We shall use examples later to demonstrate the difference between the use of s and only $s - 1$ species in the calculations. The subsequent formalization will refer to the 'all-species' mode, to simplify the discussion. The computer program DARIUS may be instructed to work in either 'all-species' or 'subcombination' mode.

(b) Computing $MCI_{(t)}$ for a random assemblage will provide positive values, as is clear from the above examples. Zero values result only if all probabilities equal 1, i.e. for very large t . Thus, the mean values themselves are meaningless without comparison to a reference basis, the expected value under the assumption of CSR. This requirement is in agreement with the uni- and bivariate approaches, and will be considered below.

(5) *Deviation of $MCI_{(t)}$ from expectation.* The expected value of $MCI_{(t)}$ under complete spatial randomness is given by the well-known Shannon entropy function. It uses the probabilities of all the possible species combinations that can be detected, i.e. the entropy is given by

$$E[MCI_{(t)}] = - \sum_{j=1}^{2^s} P[\mathbf{F}_j]_{(t)} \ln P[\mathbf{F}_j]_{(t)} \quad (5)$$

This quantity, given the vector $\mathbf{p}_{(t)}$ of probabilities of species absences at radius t , can be calculated by an algorithm which generates all the 2^s species combinations. Note that in Eq. 5 summation includes all possible species combinations, whereas in Eq. 4 summation is over species combinations actually found so that a particular combination may appear repeatedly. This distinction is expressed by different indexing: k referring to actual combinations, j to all theoretical ones. This equation, although clear in its meaning, takes much time to solve for large values of s , but there is a computationally more efficient formula for the expectation, defined as

$$E[MCI_{(t)}] = - \sum_{i=1}^s [p_{i(t)} \ln p_{i(t)} + (1 - p_{i(t)}) \ln p_{i(t)}] \quad (6)$$

(see also Erdei et al. 1994). Then, the formula

$$\Delta MCI_{(t)} = MCI_{(t)} - E[MCI_{(t)}] = \sum_{k=1}^{N'(t)} I[\mathbf{F}_k]_{(t)} / N'(t) + \sum_{i=1}^s [p_{i(t)} \ln p_{i(t)} + (1 - p_{i(t)}) \ln(1 - p_{i(t)})] \quad (7)$$

will give the *departure* of mean compositional information from its expectation, a quantity proposed for characterizing multi-species point patterns.

$\Delta MCI_{(t)}$ may be negative or positive. Most of the examples discussed aim at facilitating the interpretation of these deviations. One would expect that negative deviations correspond to a multivariate case of repulsion: there are fewer, and less diverse combinations around the individuals than expected under randomness. Positive deviation may seem to suggest aggregation: more and rare species combinations appear in the pattern. The examples will show that the interpretation is not that simple, however, and will point out

that examination of the contributions by each species should always accompany the evaluation of overall deviations.

6. *Scale dependence of $\Delta MCI_{(t)}$* . The history of pattern analysis demonstrates convincingly that the direction and magnitude of the departure from randomness is scale-dependent. The same point map can exhibit maximum negative departure for one value of t , and maximum positive deviation for a different t -value. At yet another point of scale the pattern may appear indistinguishable from a completely random arrangement. That is, for a full exploration of point patterns a series of t -values must be applied. Therefore, steps 2 - 5 are repeated for increasing values of t , and the resulting ΔMCI -values are illustrated by a graph in function of t .

7. *Significance of $\Delta MCI_{(t)}$* . Positive and negative departures from random may be simply due to stochastic variation or fluctuations without real background effects. Correct interpretation of results therefore requires simulation of a confidence envelope of the statistic for the whole range of t -values, a common practice in univariate and bivariate second order spatial analysis. Then, the multi-species point pattern may be declared significantly non-random for scale points (radii) where the actual $\Delta MCI_{(t)}$ -values lie outside the envelope. The significance level is determined by the number of $\Delta MCI_{(t)}$ -values simulated. In order to create the usual 95% confidence envelope, the minimum and maximum of 19 simulated values need to be found for each t -value (cf. Ripley 1981; Diggle 1983). We shall show that acceptance of the null-hypothesis, i.e. that a curve lies within the envelope, does not necessarily imply that absolutely random events govern pattern formation, and therefore species contributions must also be examined.

The simulation involves random placement of exactly the same number of individuals for each species as in the actual map. For a 95% envelope 19 such maps have to be created and the whole analysis repeated for each. One might argue that it is simpler to compute $N'_{(t)}$ random species combinations based on the probabilities of absences directly, without generating point maps. However, direct computations assume an infinite study area with infinite sample size, whereas real situations are always constrained. The circular plots are therefore not independent in the actual point map, so that an unusual combination is manifested several ways in the results. Thus, if we simulate combination probabilities directly, this independence disappears and a narrower, unrealistic envelope is obtained. This was tested by comparing several simulation results; we found that the discrepancy increases as the radius increases, whereas for small radii the two simulations produced

quite similar minima and maxima for the envelope. This agrees with preliminary expectations: unusual combinations have very little chance to develop for small circles.

8. *Species contributions*. One advantage of the individual-centered sampling strategy is that the role of each species in the assemblage can be evaluated. The $\Delta MCI_{(t)}$ -values can be broken up into contributions by the separate species which may also be plotted over t . Species with high negative or positive contributions are key species regarding assembly rules in the community, while species with the lowest contributions will be less important. In the evaluation of such plots one should consider that in $\Delta MCI_{(t)}$ the species are weighted according to abundance, so the effect of a rare species with non-random behaviour may be reduced by the others.

Examples

Artificial patterns were created by hand in order to show the performance of the method in easily comprehensible (albeit less realistic) situations (cf. Ripley 1987 for the univariate case). Four maps were created, each for six species dispersed in an area of 100 by 100 scaling units. To allow comparisons, the total number of individuals was similar in all maps.

1. *'Clumped' pattern*. The individuals of each species form small point clumps which are arranged more or less randomly within the region (Fig. 1a). These point clumps are separated from one another.

2. *'Regular' pattern*. Although an unrealistic situation, the regular pattern is also considered because of its ability to demonstrate different and interesting aspects of our method. Individuals occupy positions in a 17 by 17 grid with some noise added to avoid complete regularity (max 10% of the spacing distance of 5.55 units in random direction). Species identity of points is completely randomized (Fig. 1b).

3. *'Mixed' pattern*. A strongly uneven abundance distribution: species 1, 3 and 5 are 3-4 times more common than the others and they are fairly evenly distributed over the study region. Species 2, 4 and 6 form loose clumps with one another which are not isolated from the common species (Fig. 1c).

4. *'Small dense patches' pattern*. This is derived from the previous one: species 1, 3 and 5 have similar pattern with an even higher density. However, species 2, 4 and 6 form four, very small and dense clusters located in space not occupied by the common species (Fig. 1d).

Then a *simulated data set* was used, produced by a spatially explicit individual-based Monte-Carlo model of multi-species plant communities (see Czárán & Bartha 1992). Starting from a random initial community

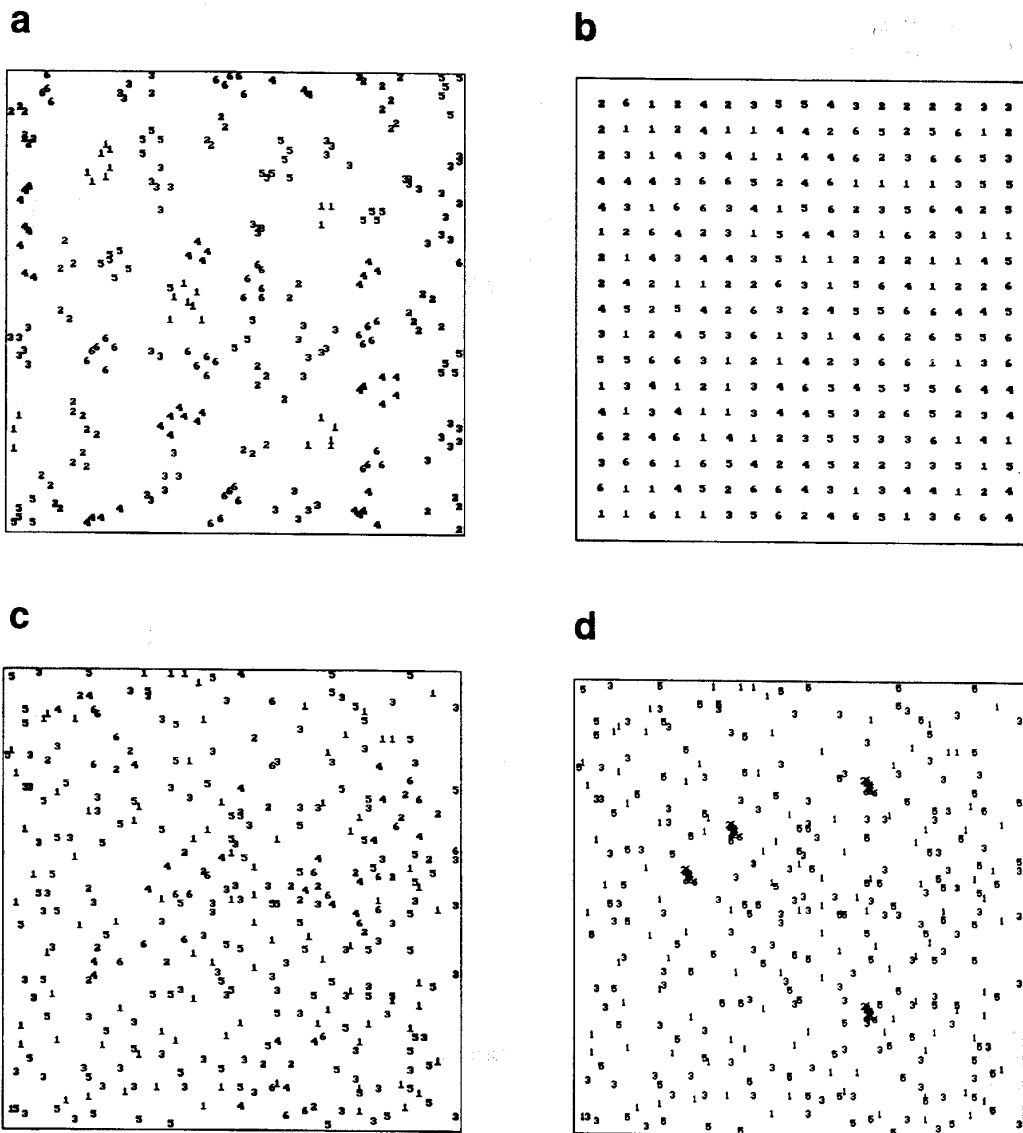


Fig. 1. Artificial point patterns referred to in the text as (a) 'clumped', (b) 'regular', (c) 'mixed' and (d) 'small dense patches'. Abundance distributions and number of points: (a) 266 [25, 55, 64, 41, 39, 42]; (b) 289 [57, 48, 42, 55, 40, 47]; (c) 339 [90, 24, 89, 23, 88, 25]; (d) 383 [113, 12, 114, 16, 112, 16]

pattern, the model generates a series of point maps representing subsequent states of the community. The spatial positions of the individuals are specified by 2D coordinates in an area of 60 by 60 units. Time is discrete, one time unit representing one generation. The model has four modules: (1) seed set, (2) dispersal, (3) seedling and adult survival and (4) competition. We introduce these briefly in turn – for details, see Czárán (1984), Czárán & Bartha (1989) and Czárán (1993).

1. Seed set. Each individual of species i of age class z produces f_{iz} seeds of which a fraction g_i will germinate, so that $\sum_z g_i f_{iz}$ seedlings will enter the competition phase. We set $g_i f_{iz} = 3$ for all 6 species in the example.

2. Dispersal. Seeds are dispersed around the parent plant with locations drawn from a 2D Gaussian distribution. The standard deviation of the distribution is the species-dependent dispersal parameter σ_i . In the simulated data set, $\sigma_i = 8.0$ for every species.

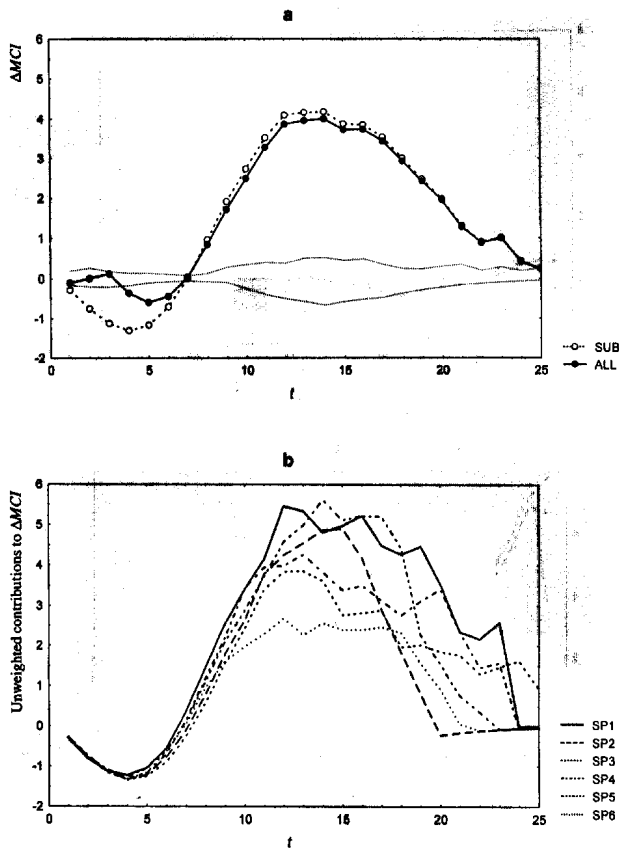


Fig. 2. Change in overall deviation (a) and species contributions to the subcombination case (b) over radius for the 'clumped' pattern (Fig. 1a). Thin lines in (a) indicate lower and upper limits of a Monte Carlo confidence envelope for 19 simulations in the subcombination mode. Note that the dotted line with empty symbols in (a) is the abundance-weighted average of the six curves in (b).

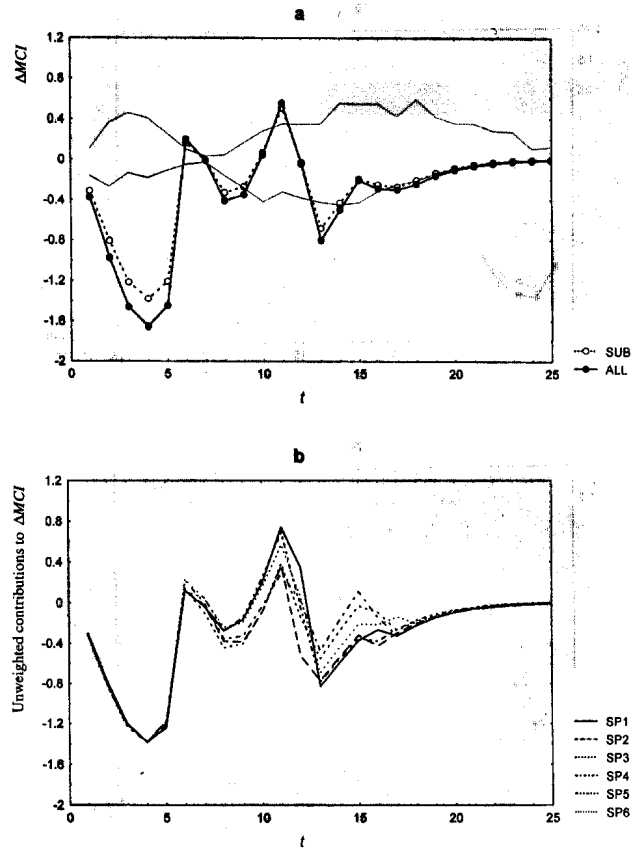


Fig. 3. Change in overall deviation (a) and species contributions to the subcombination case (b) over radius for the 'regular' pattern (Fig. 1b). The confidence envelope and abundance-weighting as in Fig. 2.

3. *Survival.* u_{iz} is the probability that an individual of species i and age z survives to age $z + 1$, assuming no competition. Our species were annuals with u_{iz} set to 0.

4. *Competition.* Seedlings experience the competitive pressure of other seedlings and adult plants which decreases their chance of surviving. The intensity of the competitive effect of an individual depends on the species, age and spatial proximity: $a_{ijz}(d)$ is the probability that a seedling of species i survives the competitive effect of a species j individual of age z at a distance d . In the simulated data set, we defined the competition functions so that intraspecific competition was strong, but interspecific competition interactions were near to zero.

Results

Artificial patterns

The deviation from random expectation of mean compositional diversity is computed for all maps using both the 'all species' and the 'subcombination' mode (Figs. 2 - 5), whereas species contributions and 95% confidence envelopes are computed and illustrated for the subcombinations only.

In the case of the *clumped* pattern (Fig. 2a), with small radii, i.e. in close vicinity of individuals, there is an obvious repulsion of other species – departure from CSR is significantly negative. For the 'subcombination' mode the negative peak is more pronounced, because the results are unaffected by the centre species. The radius of 4 units (at which the minimum of $\Delta MCI(t)$ occurs) can be interpreted as the maximum 'repulsive

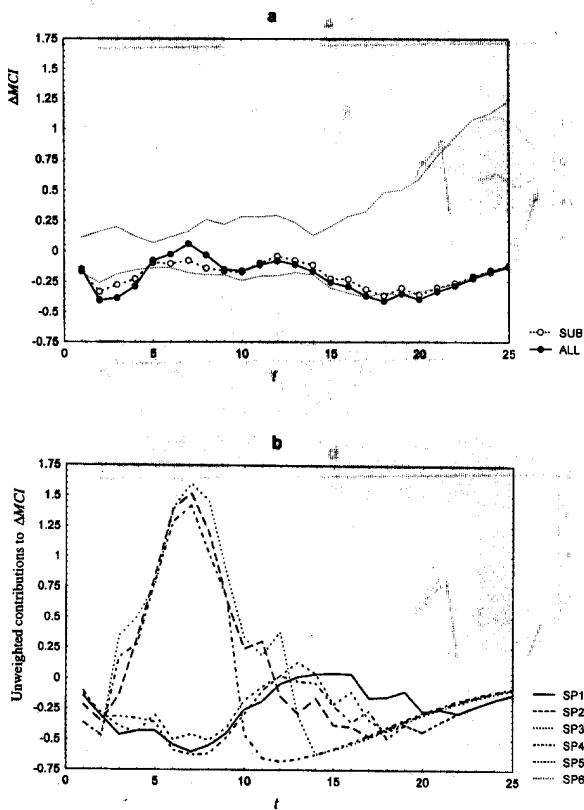


Fig. 4. Change in overall deviation (a) and species contributions to the subcombination case (b) over radius for the 'mixed' pattern (Fig. 1c). The confidence envelope and abundance-weighting as in Fig. 2.

radius' of individuals, thus reflecting average clump size in the community. For the 'all species' mode the presence of individuals of the same species as the center species is also influential. Since the species have clumped pattern, the deviations are relatively smaller. The figure shows that the difference between the two modes is diminished as the radius is increased. The pattern cannot be distinguished from random at radius 7, but up to radius 14 there is an increase of $\Delta MCI_{(t)}$ units. The maximum reflects an area at which departure of the species composition from the random case is highest. There is in fact a high chance for all species to be present in the random case. Maximum deviation therefore illustrates that this is not the case: the centre species 'attracts' certain species and repels others. At this scale point the strongest non-random species behaviour is detected. The split up of $\Delta MCI_{(t)sub}$ into species contributions (Fig. 2b) illustrates what we introduced on purpose: the species are quite similar in pattern and role in the community, that is, the overall trend of Fig. 2a does not confound species effects.

In the case of the *regular* pattern (Fig. 3a), the curve

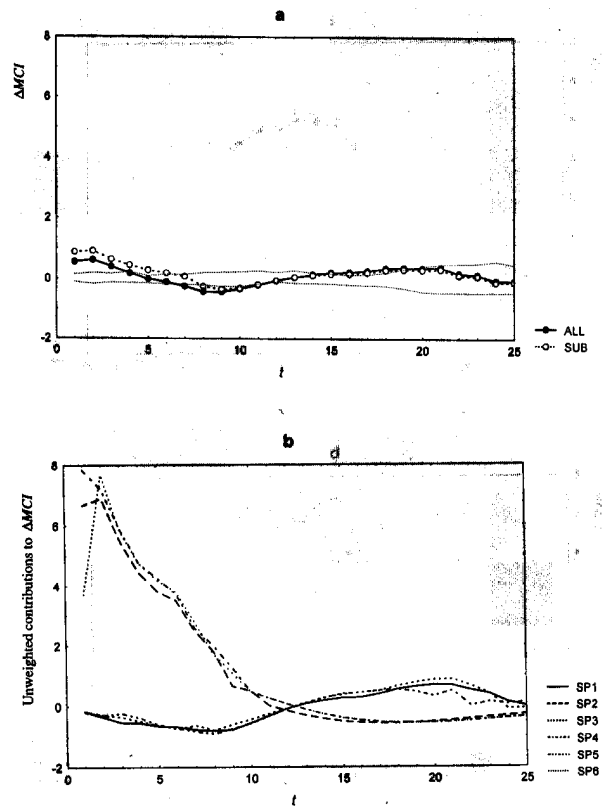


Fig. 5. Change in overall deviation (a) and species contributions to the subcombination case (b) over radius for the 'small dense patches' pattern (Fig. 1d). The confidence envelope and abundance-weighting as in Fig. 2.

of $\Delta MCI_{(t)}$ is similar for both modes. At small radii there is a strong deviation towards the negative, because there are no neighbouring points around the individuals within the spacing of ca. 5.55 units. (This is perhaps a better illustration of repulsion than the previous example.) Then, for the circle which is definitely larger than the spacing i.e. radius 6, the overall measure becomes abruptly, and 'significantly' positive, showing that the four neighbouring individuals for every centre plant give more surprising combinations than randomly expected. This effect diminishes, however, and the deviation becomes negative again. For radius 11, ca. $2 \times$ the spacing, we see another peak above the confidence envelope, which can be interpreted similarly as the peak at $t=6$. After another significant minimum, the random situation is approximated. The regular distances between consecutive minima/maxima in the curve clearly correspond to the regularity of point arrangement. Since the species were randomly assigned to the points, it is no surprise that they are similar in their contribution to the overall trend (Fig. 3b). It is for the same reason that for large sampling units the curve remains within the random envelope.

The *mixed* pattern case demonstrates that individual species contributions are an integral part of our approach. The deviation for the subcombinations remains negative for the whole range of radii (Fig. 4a), with strong fluctuations and two 'significant' troughs in the negative domain. Based on the previous examples we may assume that the first minimum corresponds to an area of negative influence around the individuals, whereas the second one reflects some regularity of spatial dispersion. The 'maximum', which lies within the confidence envelope would easily be overlooked unless the species contributions are also illustrated (Fig. 4b). This maximum is not merely by chance, because species 2, 4 and 6 do have a great tendency for a non-random behaviour: the florula around them is much richer than expected, whereas the flora around species 1, 3 and 5 is usually poorer. These effects are almost completely confounded in the overall measure.

In the *small dense patches* example, unlike in the previous ones, the overall curve 'starts' in the positive domain (Fig. 5a), reaching a significant peak at radius 2. This suggests an aggregation of species; the split up of the overall measure by species identifies the aggregated ones (Fig. 5b). The aggregation tendency is less influential for larger radii, so that the fairly even distribution of dominant species 1, 3 and 5 becomes crucial, detected as a regularity at radius 9. For further increases of t species effects are reversed (though not significant).

Simulated series of patterns

The simulation experiment was run for 20 generations. The resulting sequence of curves for the subcombination mode is shown in Fig. 6. The species were identical in all their dynamical parameters, therefore we show only the overall curves. The most obvious tendency detectable is the development of a strong positive peak shifting from $t = 5$ to $t = 8$, which attains its highest value at the 13th generation, and flattens out afterwards. The small negative peak at small sampling units is the result of the weak interspecific competition, which imposes a minimum spacing between neighbouring individuals. The positive peak is the result of short dispersal distances that yield clumped species patterns and surprising species combinations thereby. The shift of the peak with time towards larger t -values may be the result of the increase in clump size as the species spread. The decrease of the height of the positive peak after the 13th generation can be attributed to the fact that even with limited dispersal the species tend to occupy the whole plot, since neither interferes seriously with the spread of the others.

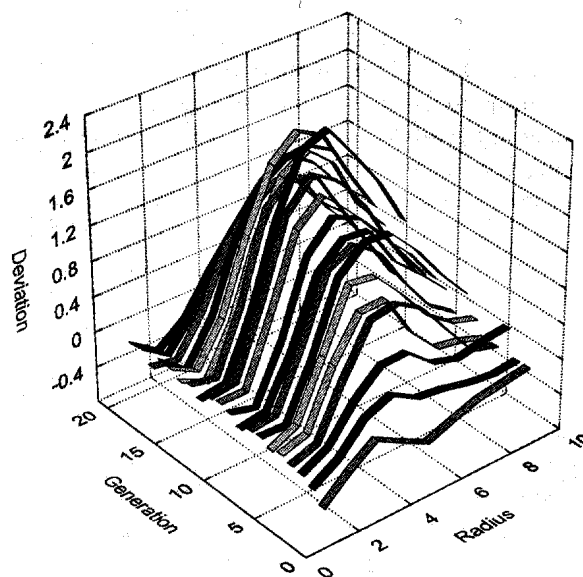


Fig. 6. Temporal change of $\Delta MCI_{(sub)}$ curves for 20 generations in the simulated pattern development experiment.

When each species has invaded the plot completely, no improbable species combinations appear in the sampling units when t is large, thus it is only the small negative peak at small t that remains on the curves, and they indicate randomness elsewhere.

Discussion

The overall measure proposed here can be used to detect an area at which the species composition of circular plots centred around plant individuals shows the highest departure from a reference basis, complete spatial randomness. The examples suggest that *negative departure* largely reflects areas of *repulsive influence* around the centre species, where other species are less likely to appear. Repulsion may appear at several scales and may be interpreted as an indication of *regularity* in the multi-species spatial pattern (Fig. 1b).

The two versions of the procedure (all-species vs subcombination mode, i.e. whether the centre species is included or excluded) differ most considerably at small scales. On the positive side differences tend to disappear, especially for large radii, and the two procedures produce very similar results for the maximum positive departure. Such departure is caused by the dominance, or at least the increased proportion, of unlikely species combinations. This is interpreted as the average radius of *maximum selective attraction* around individuals; because this is the scale point where the preference of the center species for others maximally deviates from random expectation.

The above attempt to interpret the change of overall or average departure from randomness is not without problems. One of the most important conclusions drawn from the present survey is perhaps that the single curve condensing information for all species should always be treated with caution. (This problem is analogous to the 'unicornian' property of the variance/mean ratio for single species patterns, cf. Hurlbert 1990.) Even though this curve falls within the confidence envelope, examination of species contributions may reveal that *non-random events* are concentrated around individuals of certain species. That is, when scale is sought at the community level, overall measures can be misleading if not contrasted with species-level decompositions. By the latter we do not mean conventional univariate pattern analyses, however, because a community pattern is not simply a 'sum' of species level patterns. Instead, we emphasize species contributions to the overall measure. Information theory measures (Juhász-Nagy 1976, 1993) for use in multi-species pattern analysis, seem to be suitable here. Other information statistics than the entropy measure used in this paper may also be considered in a similar approach.

The most significant difference between our method and the models developed by Juhász-Nagy is that emphasis is placed on the deviation from a null model. Therefore, the two methods are not expected to give similar results in most situations. – For example, the maximum radius for florula diversity was usually 1–3 units smaller than the area pertaining to its maximum deviation from the randomly expected. There are further differences: whereas Juhász-Nagy's models can be applied to any kind of presence/absence data, our procedure is restricted to mapped point patterns ('fully sampled patterns'). A future extension towards plot data is conceivable by incorporating permutation tests into our technique. Published applications of Juhász-Nagy's models utilize random plots located in the field or simulated by the computer (e.g. Juhász-Nagy & Podani 1983; Podani 1987) or grids and transects (e.g. Bartha & Horváth 1987). The present technique involves complete enumeration for all individuals, rather than random sampling, with the consequence that sections of the community with high abundance of individuals will be overemphasized.

This fact, that the procedure applies to point patterns only, is the most serious limitation of applicability. But, even if plants are considered point-like in space, we know that they have 'real' spatial extension (e.g. the trees in many applications of the uni and bivariate second order statistics). Some studies (Rebertus et al. 1989; Szwagrzyk 1990) treat age or some 'mass' properties of plants (e.g. trunk diameter) separately to avoid this problem. Such properties can be incorporated into our technique either by separate analyses or through a weighting procedure:

'more important' individuals may have a larger weight in the computation of $\Delta MCI_{(t)}$ than 'less important' ones.

The four artificial examples, although useful, did not exhaust even the simplest possibilities for multi-species patterns. A survey of a wider range of artificial patterns is needed to demonstrate the ability of our approach to detect scales of multi-species point patterns. Also, studies of field data are required to confirm the use of the technique in actual vegetation studies. Further investigations on dynamical situations such as the simulation case treated here, or permanent plot data series, would be necessary to extend the method to infer mechanisms of succession or habitat degradation from spatio-temporal patterns. Computer simulations would be especially useful in 'calibrating' the method for dynamical studies, as the mechanisms producing a simulated spatio-temporal pattern are known and their effect on the pattern can be directly searched for with the statistics. The method presented here may also be of interest in studies of one or two-species populations to analyse associations of different age-, size- or other classes of individuals.

Further future investigations will include, for example, the problem of edge effects, considered so important in the uni and bivariate cases. Because circles that cut the boundary cannot be considered in the analysis, available sample sizes decrease with increasing radius, and the whole analysis becomes restricted to plants located in the central area of the study region. Edge effects are most critical, and the results are most considerably biased, if the interior region is not representative of the whole study area, i.e. the pattern is not 'homogeneous' or 'stationary'. Toroidal edge correction does not work, because the opposite edges of the study region cannot be attached together (see arguments in Haase 1995, which hold even more strongly in the multi-species case). We do not see any solution in form of analytical "correction factors" at the moment. The only possible approach would be to designate a buffer zone around the study region of a width corresponding to the largest value of t to be used. In this case the assumption of homogeneity still applies, but the number of points analyzed remains constant for all values of t . Further studies of simulated and real patterns are necessary to evaluate the impact of edge effects and to examine whether the more laborious buffer zone approach produces 'superior' results.

Acknowledgements. This paper is dedicated to the memory of P. Juhász-Nagy, our late professor at Eötvös University. The authors are grateful to J. B. Wilson for stimulating discussions and to referees for their critical comments. Financial support from Hungarian National Research Grants OTKA T19364 (J.P.) and OTKA T19524 (T.C.) is gratefully acknowledged.

References

- Bartha, S. 1992. Preliminary scaling for multi-species coalitions in primary succession. *Abstr. Bot.* 16: 31-41.
- Bartha, S. & Horváth, F. 1987. Application of long transects and information theoretical functions to pattern detection I. Transects versus isodiametric sampling units. *Abstr. Bot.* 11: 9-26.
- Bartha, S., Czárán, T. & Oborny, B. 1995. Spatial constraints masking community assembly rules: a simulation study. *Folia Geobot. Phytotax.* 30: 417-482.
- Cox, G.W. 1987. Nearest neighbour relationships of overlapping circles and the dispersion pattern of desert shrubs. *J. Ecol.* 75: 193-199.
- Czárán, T. 1984. A simulation model for generating patterns of sessile populations. *Abstr. Bot.* 8: 1-13.
- Czárán, T. 1993. PATPRO: A Monte Carlo simulation program for multispecies neighbourhood competition. *Abstr. Bot.* 17: 275-281.
- Czárán, T. & Bartha, S. 1989. The effect of spatial pattern on community dynamics; a comparison of simulated and field data. *Vegetatio* 83: 229-239.
- Czárán, T. & Bartha, S. 1992. Spatiotemporal dynamic models of plant populations and communities. *Trends Ecol. Evol.* 7: 38-42.
- Diggle, P. J. 1983. *Statistical analysis of spatial point patterns*. Academic Press, London.
- Erdei, Zs., Tóthmérész, B. & Erdei, A. 1994. Linear algorithm to calculate indirect spatial statistics for completely random multi-species communities. *Tiscia* (Szeged) 28: 67-72.
- Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's K function: introduction and methods of edge correction. *J. Veg. Sci.* 6: 575-582.
- Harkness, R.D. & Isham, V. 1983. A bivariate spatial point pattern of ants' nests. *Appl. Stat.* 32: 293-303.
- Hurlbert, S.H. 1990. Spatial distribution of the montane unicorn. *Oikos* 58: 257-271.
- Juhász-Nagy, P. 1976. Spatial dependence of plant populations. I. Equivalence analysis (an outline for a new model). *Acta Bot. Acad. Sci. Hung.* 22: 61-078.
- Juhász-Nagy, P. 1993. Notes on compositional diversity. *Hydrobiologia* 249: 173-182.
- Juhász-Nagy, P. & Podani, J. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51: 129-140.
- Kenkel, N.C. 1994. Bivariate pattern analysis of jack pine - trembling aspen interaction. *Abstr. Bot.* 18: 49-55.
- Palmer, M.W. 1987. Variability in species richness within Minnesota oldfields: a use of the variance test. *Vegetatio* 70: 61-64.
- Pielou, E.C. 1972. 2^k contingency tables in ecology. *J. Theor. Biol.* 34: 337-352.
- Podani, J. 1987. Computerized sampling in vegetation studies. *Coenoses* 2: 9-18.
- Podani, J., Czárán, T. & Bartha, S. 1993. Pattern, area and diversity: the importance of spatial scale in species assemblages. *Abstr. Bot.* 17: 37-51.
- Rebertus, A.J., Williamson, G.B. & Moser, E.B. 1989. Fire-induced changes in *Quercus laevis* spatial pattern in Florida sandhills. *J. Ecol.* 77: 638-650.
- Ripley, B.D. 1981. *Spatial Statistics*. Wiley, New York, NY.
- Ripley, B.D. 1987. Spatial point pattern analysis in ecology. In: Legendre, P. & Legendre, L. (eds.) *Developments in numerical ecology*, pp. 407-429. Springer, Berlin.
- Robson, D.S. 1972. Statistical tests of significance. *J. Theor. Biol.* 34: 350-352.
- Rohlf, F.J. & Archie, J.W. 1978. Least-squares mapping using interpoint distances. *Ecology* 59: 126-132.
- Schluter, D. 1984. A variance test for detecting species associations with some example applications. *Ecology* 65: 998-1005.
- Sneath, P.H.A. & Sokal, R.R. 1973. *Numerical taxonomy*. Freeman, San Francisco, CA.
- Szwagrzyk, J. 1990. Natural regeneration of forest related to the spatial structure of trees: A study of two forest communities in Western Carpathians, southern Poland. *Vegetatio* 89: 11-22.
- Tóthmérész, B. 1994. The statistical analysis of spatial pattern in plant communities. *Coenoses* 9: 33-41.
- Williams, W.T., Lance, G.N., Webb, L.J., Tracey, J.G. & Connell, J.H. 1969. Studies in the numerical analysis of complex rain-forest communities. IV. A method for the elucidation of small-scale forest pattern. *J. Ecol.* 57: 635-654.
- Wilson, J.B. 1994. Who makes the assembly rules? *J. Veg. Sci.* 5: 275-278.
- Wilson, J.B. & Watkins, A.J. 1994. Guilds and assembly rules in lawn communities. *J. Veg. Sci.* 5: 591-600.
- Zobel, K., Zobel, M. & Peet, R.K. 1993. Change in pattern diversity during secondary succession in Estonian forests. *J. Veg. Sci.* 4: 489-498.

Received 6 December 1995;

Revision received 27 September 1996;

Accepted 20 November 1996.

App. 1. Formulae for the subcombination mode.

In the main text the formulae were derived for didactic reasons for the case when all points, regardless of the species they belong to, are considered in computing compositional information, i.e. the *all species* mode. This appendix gives formulae for the subcombination mode, that is, when individuals with the same species identity as the centre species are disregarded. Assume that the center species is c , and let $\mathbf{F}_j^{(c)}$ denote species subcombination j for the remaining $s - 1$ species. Its detection around an individual of species c has the probability of

$$P[\mathbf{F}_j^{(c)}]_{(t)} = \prod_{i=1}^{s-1} q_{i(t)} \quad (\text{A1})$$

where $q_{i(t)}$ is the same as in the main text. The information obtained by detecting this particular subcombination is

$$I[\mathbf{F}_j^{(c)}]_{(t)} = -\ln P[\mathbf{F}_j^{(c)}]_{(t)} \quad (\text{A2})$$

Let $N_{c(t)}$ denote the number of individuals of species c around which full circles of radius t can be drawn, that is $\sum_{c=1}^s N_{c(t)} = N'_{(t)}$.

Therefore, the mean information for all individuals is given by:

$$MCI_{(t)sub} = \sum_{k=1}^{N'(t)} I[\mathbf{F}_k^{(c)}] / N'_{(t)} \quad (\text{A3})$$

where $\mathbf{F}_k^{(c)}$ refers to the species subcombination found around individual k , which actually belongs to species c . The expectation for a randomly chosen individual is obtained as

$$E[MCI_{(t)sub}] = \frac{\sum_{c=1}^s \left[N_{c(t)} \sum_{j=1}^{2^{s-1}} P[\mathbf{F}_j^{(c)}]_{(t)} \ln P[\mathbf{F}_j^{(c)}]_{(t)} \right]}{N'_{(t)}} = \frac{\sum_{c=1}^s \left[N_{c(t)} \sum_{i \neq c} \left[p_{i(t)} \ln p_{i(t)} + (1 - p_{i(t)}) \ln(1 - p_{i(t)}) \right] \right]}{N'_{(t)}} \quad (\text{A4})$$

and the deviation from the expectation will be

$$\Delta MCI_{(t)sub} = \sum_{k=1}^{N'(t)} I[\mathbf{F}_k^{(c)}]_{(t)} / N'_{(t)} + \frac{\sum_{c=1}^s \left[N_{c(t)} \sum_{j=1}^{2^{s-1}} P[\mathbf{F}_j^{(c)}]_{(t)} \ln P[\mathbf{F}_j^{(c)}]_{(t)} \right]}{N'_{(t)}} \quad (\text{A5})$$

Recall that index k refers to the manifested combinations, whereas j runs over all subcombinations when species c is omitted.