

ANALYSIS OF MAPPED AND SIMULATED VEGETATION PATTERNS BY MEANS OF COMPUTERIZED SAMPLING TECHNIQUES

J. PODANI

INSTITUTE FOR ECOLOGY AND BOTANY, HUNGARIAN ACADEMY OF SCIENCES, VÁCRÁTÓT;
DEPARTMENT OF PLANT TAXONOMY AND ECOLOGY, L. EÖTVÖS UNIVERSITY, BUDAPEST, HUNGARY

(Received: 1 February 1984)

Three computer programs have been developed to perform simulated sampling based on mapped and digitized vegetation data. These programs are used in comparing simulated random patterns to real vegetation maps obtained in two different steppe communities, one from Hungary and the other from Canada. The relationships between sampling characteristics and information theoretical functions are analyzed. Possibilities for determining characteristic areas defined in terms of information theoretical functions and expected resemblance are also examined. In two simple situations it is tested whether overlap among sample plots and simplification in the representation of plant pattern strongly influence results obtained in sampling simulation experiments

Introduction

Mapped vegetation data have long been used in various fields of vegetation ecology to investigate spatial pattern, the efficiency of sampling designs and the interrelationship between population parameters. The usefulness of maps lies in the wide range of sampling techniques that can be implemented and compared automatically which would not be possible to do in a field survey. Further advantage is that artificial maps can be generated to test the performance of analytical methods. In case of point patterns, vegetation data can be easily transformed to coordinates and the whole procedure can be performed by computers. For these reasons, the applications of artificial and real maps to vegetation surveys are numerous (e.g., PENFOUND 1945, CURTIS and MCINTOSH 1950, GREIG-SMITH 1952, BORMANN 1953, COTTAM et al. 1957, O'REGAN and PALLEY 1965, O'REGAN et al. 1973, ARVANITIS and O'REGAN 1967, KULOW 1966, WILLIAMS et al. 1969, LAFRANCE 1972, NOSEK 1976, SZÖCS 1977, 1979, DIGGLE 1981, BONNICKSEN and STONE 1981) despite some disadvantages attributed to the method (e.g. edge effect, plant representation problems) discussed below.

The idea of computer simulated sampling applied to mapped vegetation emerged first in forest studies and is dated back as early as the sixties. PALLEY and O'REGAN (1961) described a technique for the evaluation of point and line sampling methods using the coordinates of individual plants stored in computer memory. Their program was modified later for sampling with circular plots (O'REGAN and PALLEY 1965). These techniques were widely used to find optimum sampling designs for estimating forest population parameters such as timber volume, basal area and frequency (ARVANITIS and O'REGAN 1967, WENSEL and JOHN 1969, O'REGAN et al. 1973). In vegetation ecology, the concept of computer assisted sampling was adopted by LAFRANCE (1972) in a pattern and ordination study of artificial communities. To my knowledge this is the only attempt to employ computer-generated, although not completely randomized samples in analyzing plot characteristics—ordination relationships.

For various purposes, Szócs (1979) developed a flexible computer program (CROCUS) offering a broad variety of sampling parameters. Randomly or systematically arranged plots of any size, orientation and three different shapes can be simulated, and distance-based sampling methods can be performed. Another program written by BONNICKSEN and STONE (1981) also contains plotless and plot sampling procedures including GREIG-SMITH's (1952) grid analysis.

As mentioned above, there are some fundamental problems associated with computerized sampling, which may introduce bias into the results. The first "defect" is due to the restriction that no overlap between sampling units and the boundary line of the mapped area is allowed. As a consequence, even though the plots are taken at random, the individuals close to the boundary will have a lower probability for being present in a random sampling unit than the others ("edge effect", O'REGAN and PALLEY 1965, WENSEL 1975). This is demonstrated by Figure 1a in which square units with orientation parallel to the edge are used for simplicity. Clearly, for any individual with distance from the boundary larger than or equal to the side length of plots, s , the probability of occurrence is proportional to s^2 . If the distance is smaller than s , say x , the probability of being sampled is lower, because all random plots are deleted whose centroid would fall into the edge strip of width $s/2$. The reduced probability is proportional to $s \cdot x$, which is illustrated by the shaded part of each square drawn around the individuals as centroids (Fig. 1a). The result is that within the edge of width s the sampling intensity is continuously decreasing towards the edge. For this reason, if quadrat

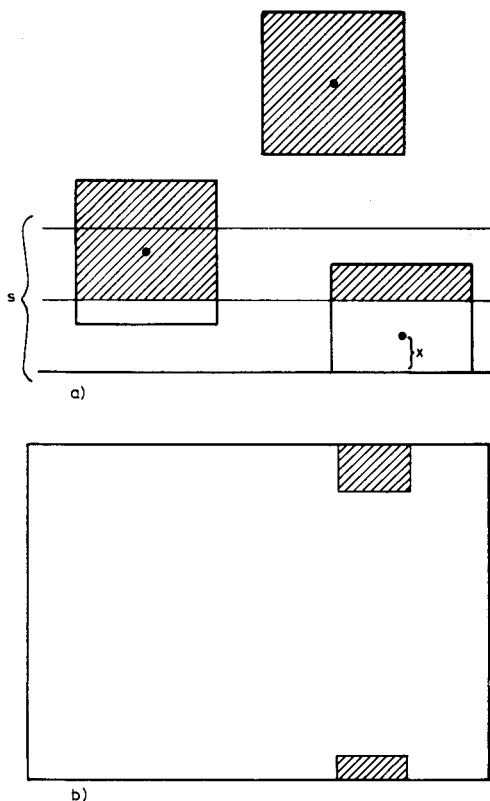


Fig. 1. Edge effect. a) Shaded areas are proportional to the probabilities for three individuals for being sampled. b) WENSEL's correction for edge effect

sampling is used in estimating, for instance, the volume of timber in a forest, the estimation will be biased. WENSEL (1975) described a technique to overcome this difficulty suggesting the inclusion of any incomplete sampling unit complemented by another fragment coming from the opposite side of the study area (Fig. 1b). Obviously, this strategy may help in estimating only some population parameters and species-individual diversity. If interest is centred upon interspecific associations, inter-plot resemblance coefficients, supra-individual diversity, etc., Wensel's technique is obviously inappropriate. It is felt, however, that in the latter cases the edge effect should not be corrected at all in a simulated study. This problem equally applies to field surveys, since random sample plots are never laid onto the boundary line of the study area. Theoretically, any field study is affected by edge effects although they become negligible as the study area increases. Thus, even though three maps used in this survey actually represent small parts of real communities, there is no reason for correcting an effect which is unavoidable in all studies. Of course, the scales of a field study and a simulated investigation are very different, but it seems immaterial concerning the recognition of principal spatial trends in vegetation analysis. The only, but very important consequence of the difference in scales is that the results from a computerized study (e.g., the size of characteristic areas) do not apply directly to the whole community from which the map was taken.

The second problem is related to the geometric representation of plant individuals on the maps. Whereas two-dimensional scattergrams may give a good approximation to the original pattern in forests, the description of grasslands and many other plant communities may be much less accurate this way. Centroids of plant patches and point clusters may well be used instead of the actual geometric shapes, but point representations are presumably not satisfactory for all purposes. However, the discrepancy between results based on point patterns and cover maps has yet to be examined. An objective of this study is to do this by comparing the results based on two different maps of the same area.

Another potential disadvantage is that random plots generated by the computer may be overlapping unless the strong restriction of taking only exclusive plots is forced on the sampling procedure (e.g., as in case of systematic sampling). If plot size or sample size increases, the area sampled with overlapping plots will also increase. Surely, this process results in increasing precision of estimates for population parameters and diversity indices [for the latter case evidence is found in NOSEK (1976)]. However, it is not yet known whether estimates of information theoretical characteristic functions (JUHÁSZ-NAGY and PODANI 1983) or expected resemblance (PODANI 1982, 1984) are biased as a result of such "oversampling". Therefore, the effect of overlapping among plots will also be analyzed in this study.

The majority of computerized sampling studies are concerned with the pattern of a single species (e.g., BONNICKSEN and STONE 1981), species-individual diversity (NOSEK 1976) and, in particular, with the performance of various sampling methods in estimating population parameters for a selected species (e.g., KULOW 1966, PALLEY and O'REGAN 1961). There is no doubt that digitized field data and simulated sampling may equally be useful in studying either small-scale community structure (WILLIAMS et al. 1969) or the influence of sampling upon the results of multivariate analyses (LAFRANCE 1972). Some other questions frequently raised in phytosociology, such as the determination of characteristic areas in plant communities or the estimation of inter-plot resemblance have not yet been investigated using the computerized sampling approach. In this study the main objective is therefore to examine how the information theoretical characteristic functions formulated by JUHÁSZ-NAGY and PODANI (1983) and expectations for four resemblance coefficients (SORENSEN index, RUSSELL-RAO index, simple matching coefficient, and Euclidean distance; see ORLÓCI 1978, and SNEATH and SOKAL 1973) are influenced by plant pattern and changes of sampling parameters. Of course, all possible aspects of this problem cannot be undertaken in a single survey, so atten-

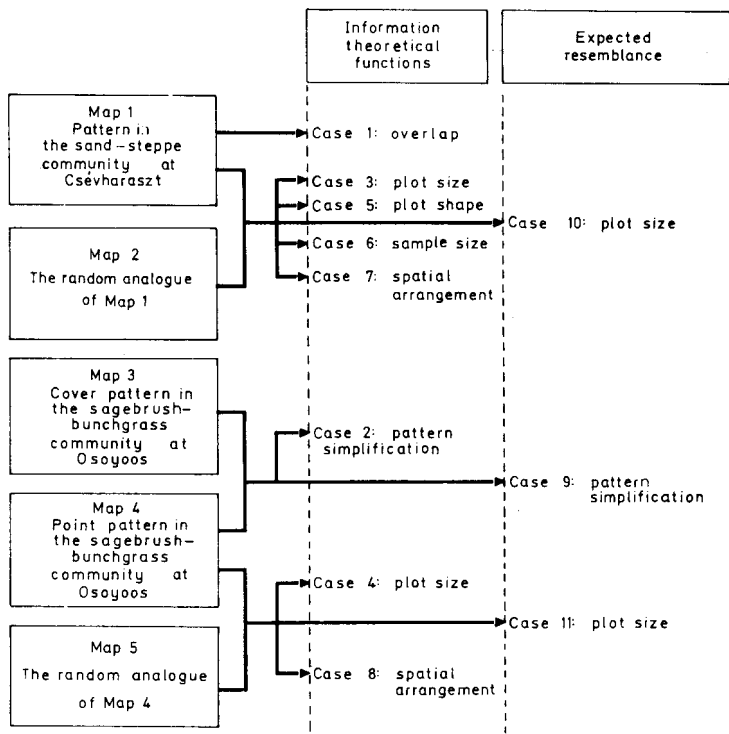


Fig. 2. Diagram illustrating the sampling experiments performed in the present study. The maps compared or analyzed, the functions calculated, and the conditions of sampling changed in each case are shown

tion will be paid only to relatively simple situations. The basic strategy is that real vegetation maps will be contrasted with simulated random patterns such that a particular sampling parameter is changed (primary process in the real space, cf. PODANI 1984). These changes are: increase of plot size or sample size, elongation of plot shape, and the transformation of systematic sampling design into a completely random arrangement of plots (sampling process). The different situations examined are summarized in Fig. 2, which demonstrates that interest will be focused on plot size changes in order to examine the potential usefulness of two different approaches to the characteristic area problem.

Material

Five maps were selected in such a way that homogeneous and heterogeneous communities, clumped and random distributional patterns are all represented to allow a comparative evaluation of results. One map was obtained from Z. Szócs, two others depicting different geometric images of the same plant assemblage were obtained by the author. In addition, two random (more precisely, pseudo-random) multi-species patterns were simulated with identical species numbers and abundances as in the real maps. These maps will be used as references to examine the departure of results from those based on the random distribution of plants.

Map 1 is the point pattern of the most abundant six species (Table 1) in a perennial sand-steppe at Csevéharaszt (Kiskunság National Park, Hungary). The $2.4 \times 10 \text{ m}^2$ area was mapped in 1973 by Z. Szócs using his photocomputational technique (Szócs 1977, 1979). Based on the photographs, the position of the centroid of each plant individual was digitized in form of two-dimensional Cartesian coordinates and committed to computer core. The digitized map was originally subjected to intensive pattern study by computerized sampling methods (Szócs 1977). The area is fairly homogeneous and most of the species exhibit clumped spatial dispersion. The random analogue of this map is Map 2.

Map 3 was taken in a sagebrush-bunchgrass community in the Okanagan Valley at Osoyoos (British Columbia, Canada) in the summer of 1981 (Fig. 3). Due to the complexity of the vegetation, the photocomputational method was not used. Instead of this, an 8×20 grid—composed of $31.25 \times 31.25 \text{ cm}^2$ cells, yielding a total area of $2.5 \times 6.25 \text{ m}^2$ —was laid down to facilitate the mapping. The positions of plant individuals and patches were carefully recorded. A total of 23 species were found (Table 2). The actual spatial extension of the individuals of 15 species was regarded as point-like. For one species (*Artemisia tripartita*) both the position of stems and the perpendicular projection of living plant parts onto the ground were drawn. Other species (*Carex filifolia*, *Antennaria rosea*) and grasses (*Agropyron spicatum*, *Bromus tectorum*, *Festuca occidentalis*, *Koeleria cristata*, *Poa sandbergi* and *Stipa occidentalis*) were plotted as patches. Since this map cannot be readily used for computerized sampling, another map (Map 4, Fig. 4) was drawn to indicate only point patterns. In this map, the position of stems and the centroids of grass patches are shown or are represented by point-clumps. Map 4 is then a simplified, ready for digitization version of Map 3. However, the simplification reduces the information content of the map, and this fact has to be considered when evaluating the results. Despite its relatively small size, this area is not homogeneous, very few species occur throughout. The number of individuals on the left side of the map is

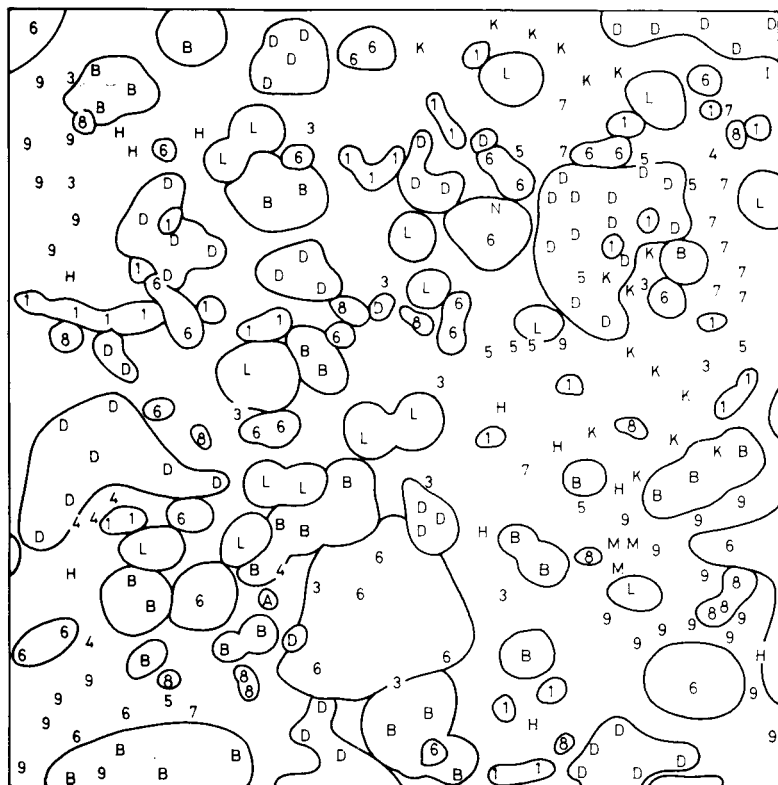


Fig. 3. Cover pattern in the sagebrush-bunchgrass community. Detail of Map 3 showing one-tenth of the total mapped area

Table 1
List of species occurring in Map 1.
Nomenclature follows Soó and JÁVORKA (1951)

No.	Name	Number of individuals
1	<i>Koeleria glauca</i>	755
2	<i>Stipa pennata</i>	147
3	<i>Festuca vaginata</i>	287
4	<i>Polygonum aviculare</i>	48
5	<i>Tragus racemosus</i>	324
6	<i>Minuartia verna</i>	50
Total:		1611

Table 2
List of species occurring in Maps 3-4.
Nomenclature follows SCOGGAN (1978)

No.	Symbol on maps	Name	Number of individuals
1	1	<i>Poa sandbergii</i>	273
2	2	<i>Erigeron pumilus</i>	29
3	3	<i>Lupinus sericeus</i>	75
4	4	<i>Plantago patagonica</i>	492
5	5	<i>Collomia linearis</i>	148
6	6	<i>Artemisia tripartita</i>	259
7	7	<i>Madia exigua</i>	337
8	8	<i>Antennaria rosea</i>	205
9	9	<i>Achillea millefolium</i>	139
10	A	<i>Bromus tectorum</i>	126
11	B	<i>Agropyron spicatum</i> var. <i>inerme</i>	166
12	C	<i>Koeleria cristata</i>	27
13	D	<i>Carex filifolia</i>	426
14	E	<i>Phlox longifolia</i>	25
15	F	<i>Stipa occidentalis</i>	40
16	G	<i>Saxifraga integrifolia</i>	31
17	H	<i>Polygonum majus</i>	103
18	I	<i>Calochortus macrocarpus</i>	5
19	J	<i>Zigadenus venenosus</i>	1
20	K	<i>Erigeron corymbosus</i>	96
21	L	<i>Festuca occidentalis</i>	51
22	M	<i>Lithospermum ruderale</i>	22
23	N	<i>Castilleja pallescens</i>	1
Total:			3077

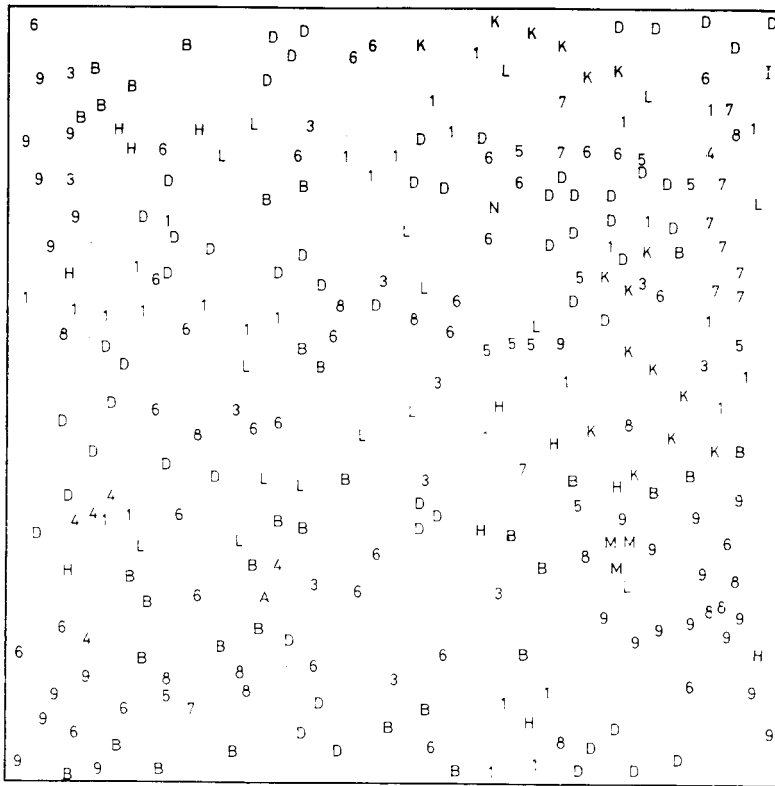


Fig. 4. Point pattern in the sagebrush - bunchgrass community. Detail of Map 4 corresponding to the one shown in Fig. 3

much higher than on the other side, due to the abundant presence of *Plantago patagonica*. Some species are almost completely restricted to the middle or the right part of the area. The random analogue of Map 4 is Map 5.

Methods

Since the existing computer programs lack some options required by the present study, it appeared practical to develop my own sampling simulator program package. Three FORTRAN programs applicable to point patterns have been written.

Program ELSAM is designed to simulate random plots of circular, rectangular or elliptical shape with random or uniform orientation. The map and the sampling units may be drawn using a HP-7221A plotter. The standard output is a species-by-sampling units matrix of counts written onto disk. Since the plots are independently selected, the possibility of their overlap is not excluded.

The sampling process (PODANI 1982, 1984) which includes systematic, restricted random and fully random sampling may be carried out by program SAMPROC. Plotting subroutine is not provided, only circular and rectangular plots and square blocks can be specified. If required, only systematic sampling is performed and in this case the output consists of a single matrix. Otherwise several matrices will result, each corresponding to one particular stage of the process.

Random pairs of plots are simulated by program EXPRES which in turn calculates the expectation of resemblance coefficients. In this program it is ensured that the paired plots are non-overlapping.

Florula diversity and related information theory functions were calculated by program INPRO described first in JUHÁSZ-NAGY and PODANI (1983). The input of this program is a species-by-quadrats matrix. The result includes the actual values for characteristic functions and the number of manifested species combinations. If required, the frequency distribution of combinations is also printed.

On Map 3 the sample plots were manually placed using a table of random numbers. In this case expected resemblance was calculated such that random pairs of plots were selected from the data matrix. In order to ensure independence among these pairs, every plot was considered once. Expectations for four binary coefficients (SORENSEN index, RUSSELL-RAO index, simple matching coefficient, and Euclidean distance) were calculated for all types of maps.

Results

Information theoretical functions vs sampling characteristics

Case 1: The effect of overlap among plots

This problem was examined using the Csévharaszt map (Map 1). Program ELSAM generated 250 random and potentially overlapping plots at each of the ten plot sizes previously selected. The plot shape was square, the sizes ranged from $10 \cdot 10 \text{ cm}^2$ to $60 \cdot 60 \text{ cm}^2$. Non-overlapping plots with similar sizes were placed in systematic arrangement by program SAMPROC. The sample size was 64 with the largest plots forming a continuous grid. The relationships of characteristic functions and plot size are given in the diagrams in Fig. 5. It is seen that the two results are in very close agreement, the corresponding curves are similar in shape. Local distinctiveness and florula diversity are high within the interval $[0.1, 0.2]$ in both cases. The actual maxima occurred at $40 \cdot 40 \text{ cm}^2$ when the sampling was systematic, at $37.5 \cdot 37.5 \text{ cm}^2$ when the plots were randomly arranged. Nevertheless, the peaks are not striking and the differences in maximum areas indicated appear negligible. Further correspondence between the two results is that associatum shows less obvious trend than the other two functions. A relatively high level is maintained within a broad range of plot sizes and no significant peaks can be seen. These observations suggest that overlapping plots produce no artifacts and can be used to estimate information theoretical and probably many other characteristic functions.

Case 2: The effect of pattern simplification

The discrepancy between results based on two geometric images of the same area was examined using the Osoyoos maps (Maps 3-4). The plot shape was circular, with radii ranging from 0 cm to 75 cm. The sample size was 250 both in the simulated and the manual sampling. Figure 6a shows the change of the number of species combinations, Figure 6b illustrates the characteristic functions. Surprisingly, the results are more similar than expected. The position of peaks is invariant, the maximum of local distinctiveness and associatum occurs at radius of 25 cm in both maps. Florula diversity does not exhibit a peaked effect, there is a characteristic interval between 15 and 25 cm, rather than a particular well-indicated area. Similarly, the number of species combinations (NSC) is very high at several plot sizes if the cover map is used. For the point pattern, however, NSC has a well-defined maximum. Another striking difference between the results is that the actual values of characteristic functions at

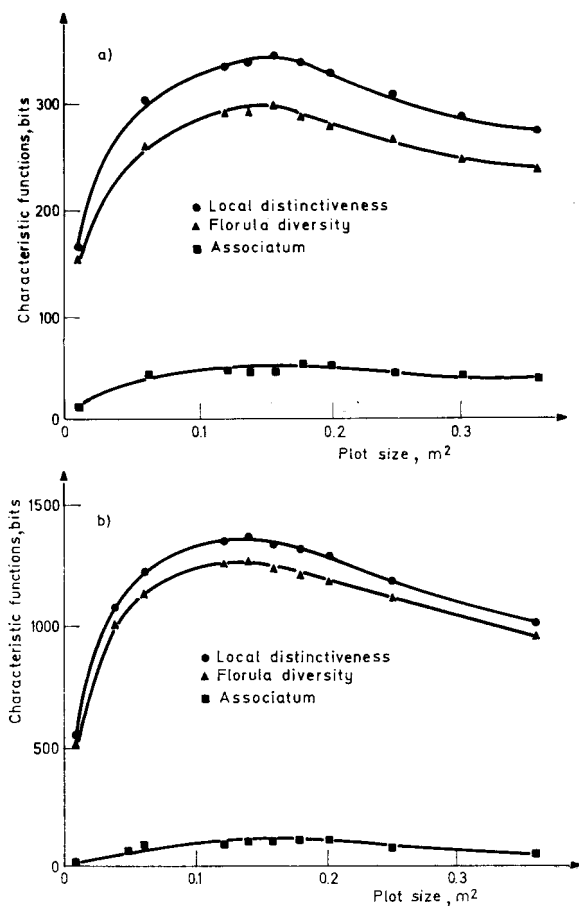


Fig. 5. Case 1: The effect of the size and arrangement of plots on characteristic functions in the sand-steppe community at Csévharaszt. a) Systematic sampling (no overlap allowed). b) Random sampling (plots may overlap). Note that differences in magnitude are due to differences in sample size

small plot sizes tend to be higher in case of the cover map. These are the only consequences of the divergence between the two contrasting pattern representations of the same area. The congruence of results is unexpected and probably due to the fact that 15 of the 23 species are represented by points in both maps. This number is large enough to ensure invariability in the size of characteristic areas. Of course, this is not to say that the present findings have unrestricted validity, since the distortion introduced by simplification to point patterns depends on the number of species and their individuals with patchy spatial appearance. Nevertheless, in case of the Osoyoos map simulated sampling may well be used instead of the more laborious manual sampling.

Cases 3-4: The determination of characteristic areas

The range of plot sizes by which the Csévharaszt map was sampled in Case 1 was fairly narrow to facilitate a satisfactorily large sample size in systematic sampling. Therefore, the map was re-sampled by a series of circular plots with radii ranging from 5 to 75 cm. The

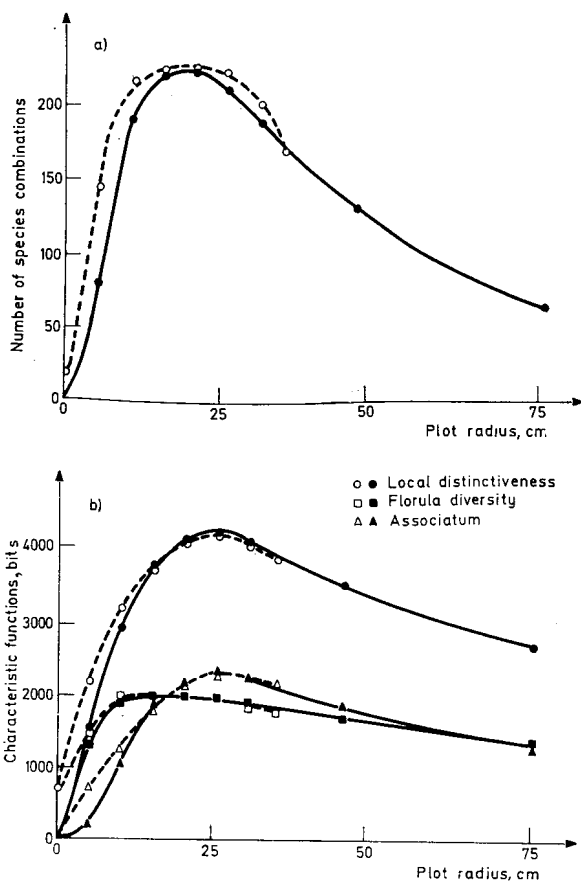


Fig. 6. Case 2: The effect of plot size and pattern simplification on the number of species combinations (a) and characteristic functions (b) in the sagebrush - bunchgrass community. Dotted line and open symbols: manual sampling of the cover map, solid line and full symbols: simulated sampling

sample size (250) was not changed. Preliminary results suggested that the critical range of characteristic areas was between circles with radii 15 and 25 cm. To test whether the peaks occurred consistently at the same point, three replicate samples were taken at plot radii 15, 20 and 25 cm. The random point pattern corresponding to the Csévharaszt map was also sampled with similar plot and sample sizes. The results are shown in Figs 7-8.

Figure 9 illustrates the results for the actual Osoyoos map and the corresponding randomly generated pattern. In these cases sample size was 300, plot radius ranged from 5 to 75 cm.

Figures 7-9 demonstrate that local distinctiveness, florula diversity, associatum and NSC first rapidly increase and later slowly decrease with increasing plot size, regardless the type of the spatial distribution of species. This observation supports the findings by JUHÁSZ-NAGY (1967, 1976) and JUHÁSZ-NAGY and PODANI (1982) who made similar surveys using field data.

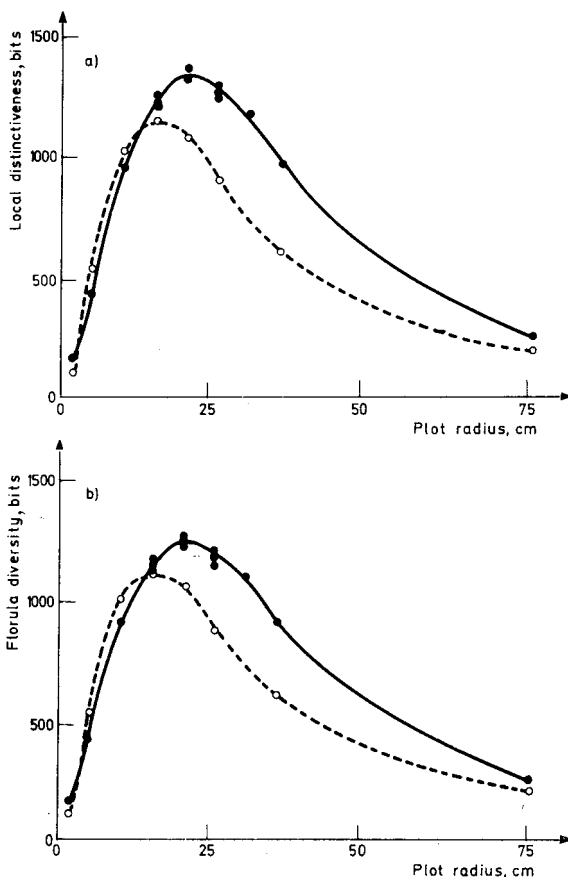


Fig. 7. Case 3: Dependence of local distinctiveness (a) and florula diversity (b) on plot size and spatial pattern in the sand-steppe community. Dotted line and open symbols: random pattern. solid line and full symbols: real map. Curves fitted by eye

The replicate samples from the Csévharaszt map indicate that there is a unique maximum for local distinctiveness and florula diversity at plot radius of 25 cm, whereas the maximum of NSC occurs at 20 cm. For the randomly dispersed population the corresponding areas are smaller. In practice, at small plot sizes there is no substantial difference between the random and the actual population. However, at about the size of characteristic areas the diversity of the actual vegetation is much higher than that of the random community, and the deviation from randomness apparently increases as the plots are further enlarged. Later the difference tends to disappear suggesting that in a homogeneous vegetation there is a fairly large plot size at which the type of spatial distribution of species becomes immaterial.

The change of associatum versus plot size (Fig. 8a) is less obvious. Although both curves have double peaks, this is not sufficient evidence for bimodality. One could easily select points from the replicate triplets such that the maximum occurs at either of 15, 20 or 25 cm plot radius. Therefore, Fig. 8a rather suggests that the estimation of associatum is more influenced by random fluctuations than are the other functions. Perhaps there is no

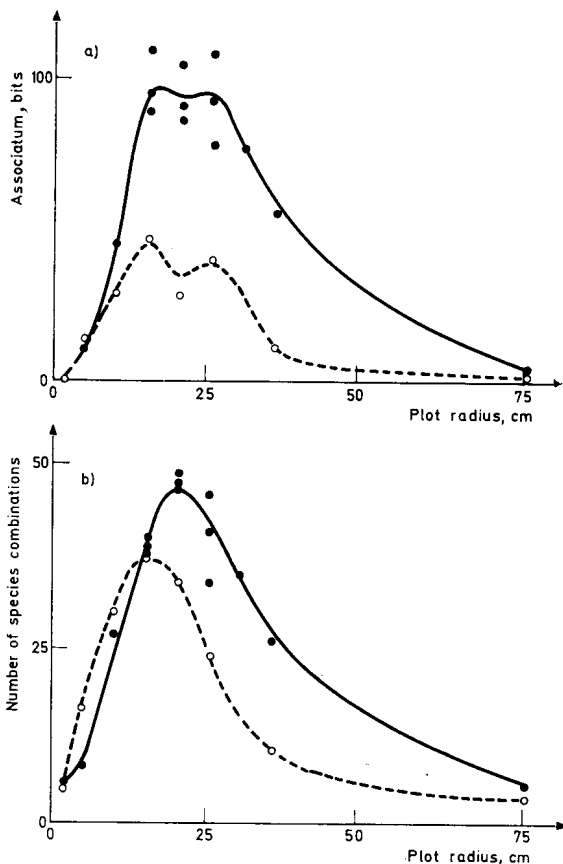


Fig. 8. Case 3: Dependence of associatum (a) and the number of species combinations (b) on plot size and spatial pattern in the sand-steppe community. Dotted line and open symbols: random pattern; solid line and full symbols: real map. Curves fitted by eye

particular maximum area but a relatively wide range of plot sizes (characteristic interval) with consistently high associatum values. The other Csévharaszt results (Case 1, see Fig. 5) also support this observation.

The Osoyoos results (Fig. 9) are similar to the Csévharaszt ones in many respects. Local distinctiveness and NSC both have a clear maximum at 25 cm, and the characteristic areas are consistently smaller in the random "community". However, the decrease of characteristic functions beyond the maxima is very slow, and the deviation from the random situation rapidly increases with plot size, at least within the range of plot sizes used. Contrary to the Csévharaszt results, the curve of associatum has a distinct peak and it is the florula diversity for which no well-defined maximum area was found. In the random population the number of species combinations reaches a higher maximum than in case of the actual community.

It may be concluded that differences in the above findings are indicating differences between the structure of the communities being analyzed. The post-maximum slope of characteristic curves and their deviation from the random case are good indicators of homogeneity (or heterogeneity). Well-defined maxima may be expected for local distinctiveness and NSC

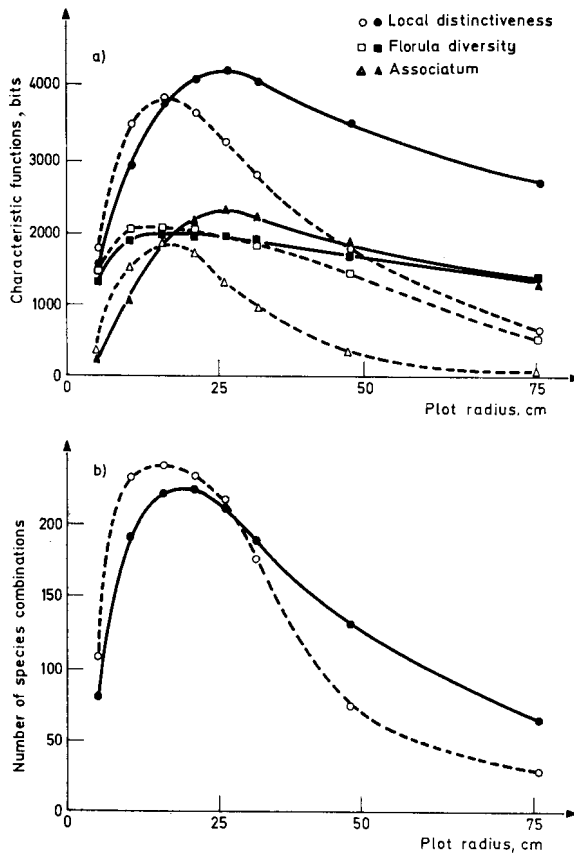


Fig. 9. Case 4: Dependence of characteristic functions (a) and the number of species combinations (b) on plot size and spatial pattern in the sagebrush – bunchgrass community. Dotted line and open symbols: random pattern; solid line and full symbols: real map

regardless the type of species pattern. Florula diversity, however, is more influenced by spatial factors and the maximum is closely approximated at various plot sizes in heterogeneous or species rich communities. The results also indicate that associatum is likely to be influenced by random fluctuations if the number of species is low. Increase in species number diminishes this effect, as the Osoyoos results clearly demonstrate.

Case 5: The effect of plot shape

The change of characteristic functions over plot size in the sand-steppe community was examined using two different isodiametric shapes: square (Fig. 5) and circular (Figs 7–8). The maxima for local distinctiveness and florula diversity occurred at $37.5 \cdot 37.5 \text{ cm}^2$ (0.1406 m^2) and at plot radius 20 cm, which corresponds to an area of 0.1256 m^2 . The difference between these two maxima seems negligible, because the area increments for the circular plots were much larger than those for square units.

The spatial process representing the elongation of plots with a constant area was examined using the Csévharaszt map. The plot size selected was 0.122 m^2 , with the following

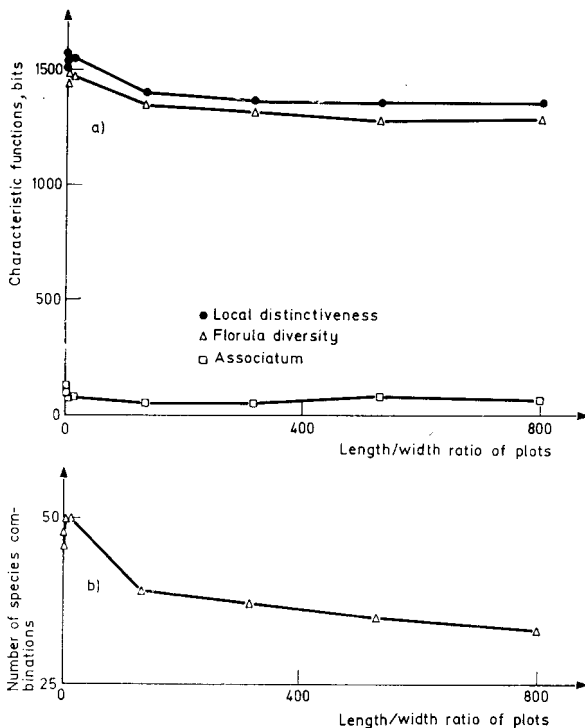


Fig. 10. Case 5: The influence of plot shape on characteristic functions (a) and the number of species combinations (b) in the sand-steppe community

series of length/width ratios: 1 : 1, 2 : 1, 3 : 1, 5 : 1, 12 : 1, 135 : 1, 325 : 1, 543 : 1, 818 : 1. The last terms of this sequence represent extremely long shapes which give a close approximation to the linear plots. As GREIG-SMITH (1964) pointed out in such a series the side length of plots increases as the ratio increases, introducing uncontrolled edge effects into the results. The only situation in which this problem could be ignored is if all the plant individuals were point-like, but this condition is never satisfied in natural vegetation.

The effect of the elongation of sampling units on the characteristic functions is demonstrated by Fig. 10. It is seen that plots with low length/width ratio result in a more diverse sample than the long units, even if fairly homogeneous point patterns are examined. Although no other experiments were made, it seems obvious that this effect is more emphatic in a heterogeneous vegetation, since longer plots are more likely to include species that are spatially isolated. However, differences between shapes that are commonly used in the phytosociological practice (square and short rectangles) do not appear significant.

Case 6: The effect of sample size

Since absolute values of characteristic functions are directly proportional to the sample size, standardization is required. Local distinctiveness is normalized if divided by its possible maximum. Florula evenness (JUHÁSZ-NAGY and PODANI 1983) may be used instead of florula diversity. The knowledge of these standardized measures is sufficient to predict the behavior of normalized associatum, which would be very difficult to calculate directly. NSC requires no standardization.

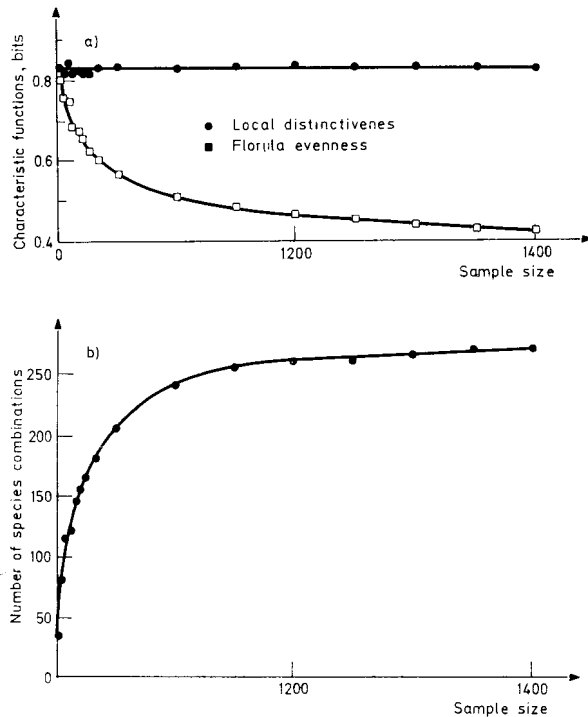


Fig. 11. Case 6: The effect of sample size on characteristic functions (a) and the number of species combinations (b) in the sand-steppe community

Figure 11 indicates that in case of local distinctiveness, sample size is a matter of statistical estimation. A small number of plots is obviously inappropriate to yield reliable estimates, but beyond a critical number, the results are not fluctuating significantly. Florula evenness, however, decreases monotonically, and therefore the normalized associatum increases as the number of plots increases. Its principal reason is that the maximum number of species combinations, which may be potentially manifested in the sample at the given plot size, is not approached even by more than two thousand plots (Fig. 11b). As a consequence, florula evenness estimates coming from different surveys can be compared only if the sample sizes are equal. It is noted that in species/individual diversity studies a similar relationship was found between evenness and sample size by other authors (see PIELOU 1975).

Cases 7-8: The effect of the spatial arrangement of plots

The Osoyoos and Csévharaszt maps were subjected to the sampling process which involved systematic, stratified random and random sampling with many intermediate stages. The sample size was considerably large, 250 and 262, respectively. Two plot sizes ($20 \cdot 20$ and $5 \cdot 5$ cm²) were used in both cases. The results are summarized in Fig. 12.

It is seen that the use of small compartments gives deviating estimates for the characteristic functions. As the design approaches full randomization, this fluctuation tends to disappear, and the last two stages of the process yield almost identical results. In case of the Csévharaszt community, systematic sampling gave the highest estimates for local distinctiveness and florula diversity, but no similar tendency is observed for the Osoyoos data. The

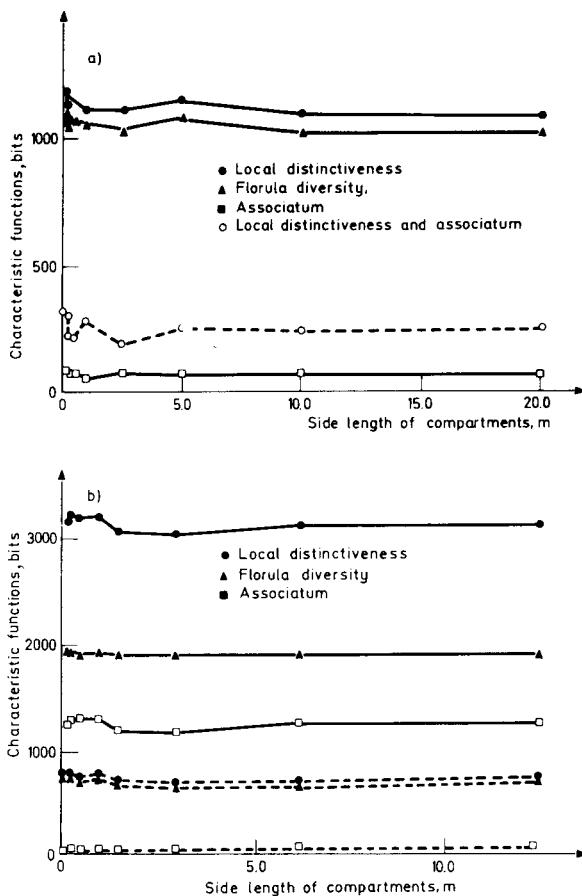


Fig. 12. Case 7: The effect of the spatial arrangement of plots on characteristic functions in the sand-steppe community (a). Case 8: The same effect in the sagebrush - bunchgrass community (b). Solid line: 20×20 cm² plots, dotted line: 5×5 cm² plots

results support the view that there is a direct relationship between the arrangement of sample plots and the precision of estimates. Systematic and stratified random sampling is not suggested for use when the goal of the survey is statistical estimation. The starting position of the grid and the size of compartments are strong influential factors. For these reasons, randomized sampling appears a more straightforward design in estimating these characteristic functions.

Expected resemblance vs plot size

Case 9: The effect of pattern simplification

It was shown in Case 2 that the simplification of Osoyoos map into a point pattern induced no significant changes in the information theoretical characteristic functions. According to Fig. 13, the expectations for simple matching coefficient and binary distance are also

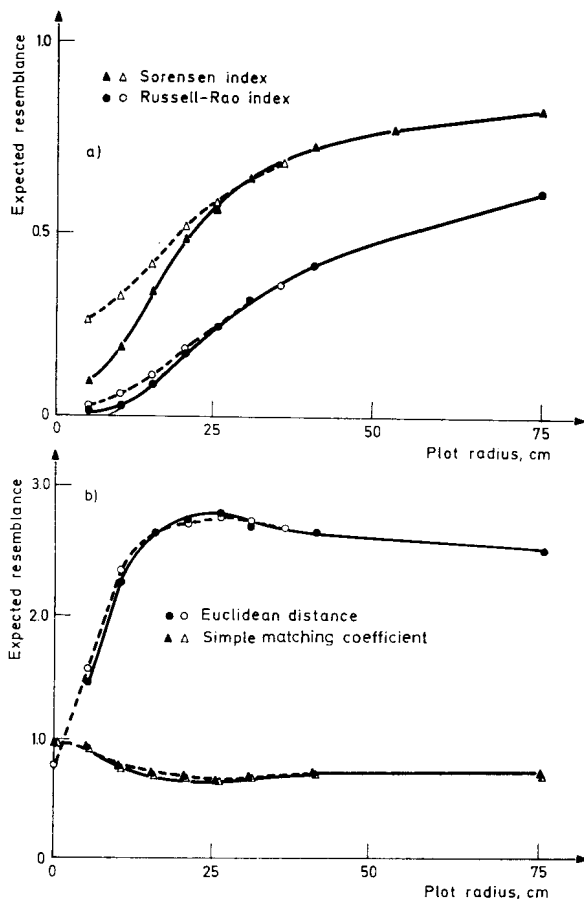


Fig. 13. Case 9: The effect of plot size and pattern simplification on expected resemblance in the sagebrush - bunchgrass community. Dotted line and open symbols: manual sampling of the cover map; solid line and full symbols: simulated sampling

little affected. The only considerable distortion occurs in case of coefficients that do not take "negative" matches into consideration (SORENSEN index and RUSSEL-RAO coefficient). However, the deviation between the corresponding curves decreases as plot size is increased. At about the area of maximum local distinctiveness the results are almost identical. The conclusion is that computerized sampling of point patterns leads to underestimation of the true expectation for these indices at small plot sizes. This fact should be considered in evaluating the results.

Cases 10-11: Expected resemblance and characteristic areas

From a pragmatic point of view it may be an important question whether expected resemblance is applicable to determine characteristic areas of plant communities. In terms of expected resemblance, a theoretical maximum area is defined as a plot size at which the expectation for inter-plot distance (or dissimilarity) is maximum, so that the sample hetero-

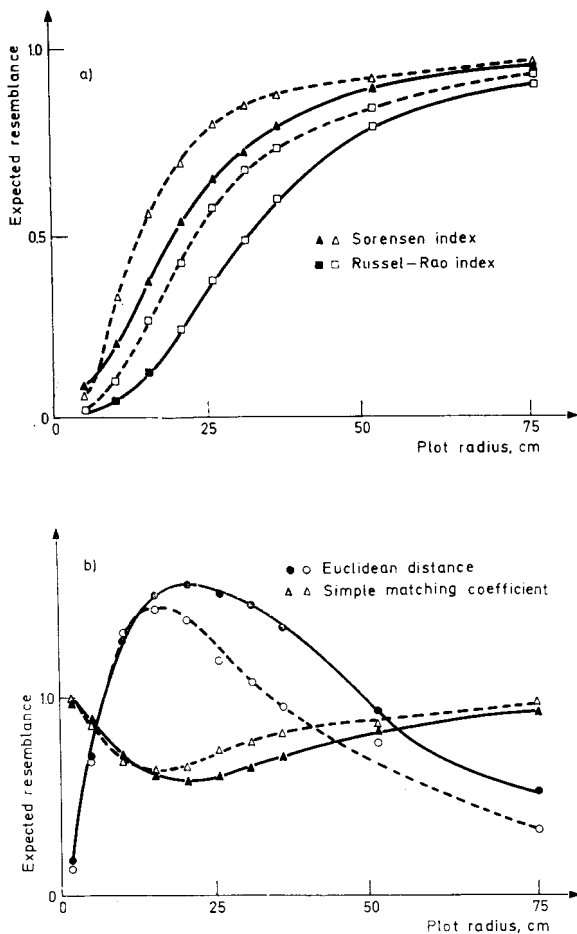


Fig. 14. Case 10: Dependence of expected resemblance on plot size and spatial pattern in the sand-steppe community. Dotted line and open symbols: random pattern, solid line and full symbols: real map

genity is also maximum. By definition, at minimum area the distance becomes zero. It is easy to see that in the binary case, as in this study, this minimum area equals the one defined in terms of florula diversity (i.e., distance may be zero only if all sample plots are similar in floristic composition). The maximum areas determined using the two approaches may be different, however, since maximum florula diversity does not necessarily imply maximum expected distance.

As already seen in Case 9, and also in Figs 14a and 15a, two binary coefficients (SORENSEN index and RUSSELL-RAO coefficient) monotonically increase over plot size. This is in fact a consequence of the exclusion of "negative" matches. Thus, only minimum area can be unequivocally determined, i.e., the plot size at which expected similarity is unity. It would be difficult, on the other hand, to recognize a point or even an interval which would correspond to the most diverse sample. The determination of maximum area by measuring the slope of

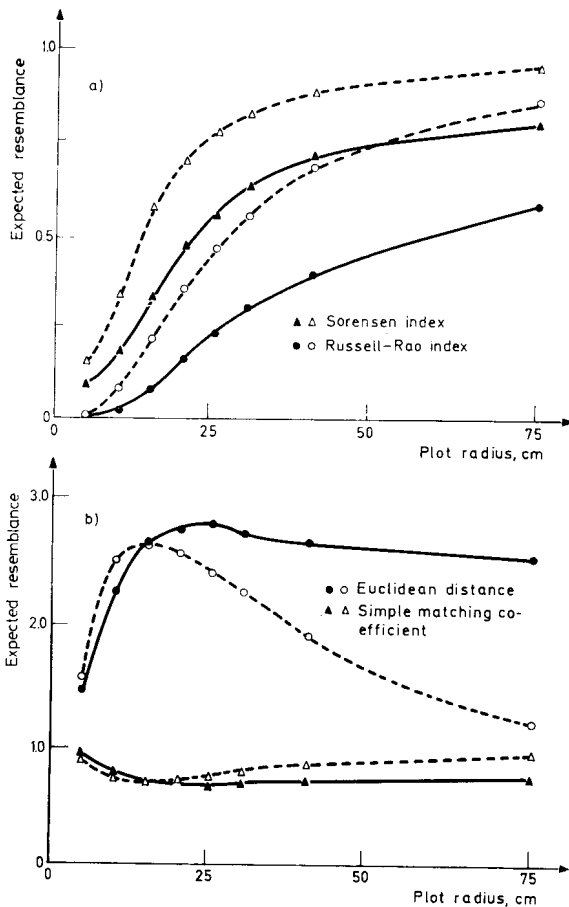


Fig. 15. Case 11: Dependence of expected resemblance on plot size and spatial pattern in the sagebrush - bunchgrass community. Dotted line and open symbols: random pattern; solid line and full symbols: real map

the curve seems rather arbitrary. As with the species/area curves, the selection of a point where the curve becomes flattened would be even more an arbitrary step.

The concept of distance-related resemblance appears more meaningful in seeking maximum areas. For Euclidean distance and simple matching coefficient, which can be calculated from one another, extreme values are indicated both in the real and the random maps (Figs 14b and 15b). Furthermore, there is a good agreement between the results obtained by these measures and the information theory functions. In the sagebrush community and its random version, the two different approaches indicate the same maximum areas. In the Csévharaszt community the area of maximum expected distance equals the plot size where maximum NSC was found. It seems then that the distance-based approach may replace the other without much loss of information. The significance of this finding lies in the fact that the estimation of florula diversity usually requires very large sample size.

The effect of spatial pattern on the resemblance of plots is also demonstrated by Figs 14-15. As expected, the sample plots from random communities are more similar to one another

than those from the real maps, unless the plots are too small. In a homogeneous community, the deviation from randomness is not significant and tends to decrease at large plot sizes (Fig. 14). In the heterogeneous community of Osoyoos, this trend cannot be observed within the range of plot sizes used (Fig. 15). Other difference is that the peak at 25 cm is not distinct and the decrease in expected distance is very slow. This is yet another indication of heterogeneity.

Concluding remarks

The findings of the present study suggest that simulated sampling is in many cases an efficient substitute for the more laborious field sampling. The analysis of vegetation maps by means of computer programs is useful when the relationships between characteristic functions and plant pattern are investigated under different conditions of sampling.

It was found that the random arrangement of potentially overlapping plots is not inferior to systematic sampling. Moreover, it is expected to give more reliable estimates for the characteristic functions since there is no upper limit of sample size. The edge effect is considered an unavoidable concomitant of plot sampling and its correction is not recommended when characteristic areas are to be determined. The sizes of characteristic areas (namely the maximum area in terms of information theory functions and expected resemblance) were little affected by the simplification of an actual vegetation map, showing horizontal projection of several species, into a point pattern. Estimates for characteristic functions and expected resemblance based on the contrasted pattern representations differed significantly only at very small plot sizes that are not used in practice anyway. One is warned, however, that simplification of plant pattern cannot always be done without much loss of information, the success is case-dependent.

The experiments with characteristic functions justified the results deriving from earlier field studies. Florula diversity, associatum, local distinctiveness and the number of species combinations in the sample are dependent on plot size and the spatial pattern of species. In a homogeneous vegetation, these functions have maxima at a certain plot size or within an interval of sizes, indicating characteristic areas at which the community is found in its most diverse development. If the sampled area is heterogeneous, characteristic areas or intervals may also be indicated as the Osoyoos results demonstrate. In theory, these functions may be best used to determine optimum plot size in presence/absence-based multivariate studies. However, the practical usefulness of this approach is limited since the estimation of florula diversity requires an unrealistically large sample size unless the number of species is very low. The problem is further complicated by the fact that the characteristic areas may vary with community types.

One alternative to the information theory functions is expected resemblance which can be derived from relatively small samples. The results obtained

by computer simulated sampling showed that distance-related binary coefficients are meaningful in seeking characteristic areas. There was a good agreement with the information theory approach although in the heterogeneous case the indication of maximum area was less obvious. This observation suggests that florula diversity is closely related to distance, whereas the relationship between similarity coefficients ignoring "negative" matches and diversity is unclear.

Characteristic areas are found in randomly simulated communities analyzed using both approaches. Thus, the existence of characteristic areas is a mathematical necessity, and the actual sizes are determined by the number of "species" and "individuals" per unit area (density) in the simulated community. In addition to species number and abundance, characteristic areas of real communities are of course also influenced by plant pattern, and the departure from the random situation offers a possibility for evaluating community structure. In the examples analyzed the influence of plant pattern was clearly demonstrated: the maximum areas for expected distance, florula diversity and related functions were always larger in the real maps than in the random analogues. Further simulation experiments are necessary to reveal the relative importance of species number, abundance and pattern in determining the characteristic areas of a community.

Parameters of sampling other than plot size were also studied. Plot shape appeared a less influential factor than size. The change of characteristic functions was little affected at the beginning of the elongation process suggesting that the choice between square and shortly rectangular quadrats is not crucial. In theory, however, isodiametric plots should be preferred for typification since this shape is the less sensitive to vegetational heterogeneities. Very long plots cannot be recommended in general, but in special cases (when the plants are point-like and the objective is estimation of population parameters for a single species) they can prove more efficient than the others.

The increase of sample size is usually the best way to increase the precision of estimates. The only exceptions are some specific derived variables normalized by theoretical maxima that cannot be observed at small sampled areas (small plot or sample sizes). Florula evenness is an example, which is expected to decrease until all species combinations potentially manifested in the sample are captured by sampling. Since this number is usually very large, the change of florula evenness over sample size is of no practical value. Results derived from different surveys can be compared only if the sample sizes were equal.

Experiments with the sampling process demonstrated that the precision of estimates is influenced by the spatial arrangement of plots. Systematic designs may result in unreliable estimates, but the fluctuations tend to disappear as the sampling approaches full randomization.

ACKNOWLEDGEMENTS

The present paper is based on portions of a PhD thesis prepared at the Department of Plant Sciences, University of Western Ontario, London, Canada. The author is grateful to his supervisor, Prof. L. ORLÓCI, for guidance and help, and to Dr. Z. SZŐCS (Vácrátót) who generously made available a vegetation map. I wish to express my thanks to Prof. P. JUHÁSZ-NAGY (Budapest) and many other colleagues and fellow graduate students for discussions. Financial support in the form of a graduate scholarship from the province of Ontario is gratefully acknowledged. The paper could not have been written without Prof. ORLÓCI's NSERC grant which covered all expenses of my research project.

REFERENCES

- Arvanitis, L. G., O'Regan, W. G. (1967): Computer simulation and economic efficiency in forest sampling. *Hilgardia* **38**: 133-164.
- Bonnicksen, T. M., Stone, E. C. (1981): The giant sequoia-mixed conifer forest community characterized through pattern analysis as a mosaic of aggregations. *Forest Ecol. Management* **3**: 307-328.
- Bormann, F. H. (1953): The statistical efficiency of sample plot size and shape in forest ecology. *Ecology* **34**: 474-487.
- Cottam, G., Curtis, J. T., Catana, A. J. (1957): Some sampling characteristics of a series of aggregated populations. *Ecology* **38**: 610-622.
- Curtis, J. T., McIntosh, R. P. (1950): The interrelationships of certain analytic and synthetic phytosociological characters. *Ecology* **31**: 434-455.
- Diggle, P. J. (1981): Binary mosaics and the spatial pattern of heather. *Biometrics* **37**: 531-539.
- Greig-Smith, P. (1952): The use of random and contiguous quadrats in the study of the structure of plant communities. *Ann. Bot. N. S.* **16**: 293-316.
- Greig-Smith, P. (1964): *Quantitative Plant Ecology*. Butterworths, London.
- Juhász-Nagy, P. (1967): On some "characteristic areas" of plant community stands. *Proc. Colloq. Inf. Theory, Bolyai Math. Soc., Debrecen*. pp. 269-282.
- Juhász-Nagy, P. (1976): Spatial dependence of plant populations. I. Equivalence analysis (an outline for a new model). *Acta Bot. Acad. Sci. Hung.* **22**: 61-78.
- Juhász-Nagy, P., Podani, J. (1983): Information theory methods for the study of spatial processes and succession. *Vegetatio* **51**: 129-140.
- Kulow, D. L. (1966): Comparison of forest sampling designs. *J. Forestry* **64**: 469-474.
- LaFrance, C. R. (1972): Sampling and ordination characteristics of computer-simulated individualistic communities. *Ecology* **53**: 387-397.
- Nosek, J. N. (1976): Comparative analysis of some diversity functions under different conditions of sampling in sandy meadow. *Acta Bot. Acad. Sci. Hung.* **22**: 415-436.
- O'Regan, W. G., Palley, M. N. (1965): A computer technique for the study of forest sampling methods. *Forest Sci.* **11**: 99-114.
- O'Regan, W. G., Seegrist, R. W., Hubbard, R. L. (1973): Computer simulation and vegetation sampling. *J. Wildl. Management* **37**: 217-222.
- Orlóci, L. (1978): *Multivariate Analysis in Vegetation Research*. Junk, 2nd ed. The Hague.
- Palley, M. N., O'Regan, W. G. (1961): A computer technique for the study of forest sampling methods. I. Point sampling compared with line sampling. *Forest Sci.* **7**: 282-294.
- Penfound, W. T. (1945): A study of phytosociological relationships by means of aggregations of colored cards. *Ecology* **26**: 38-57.
- Pielou, E. C. (1975): *Ecological Diversity*. Wiley, New York.
- Podani, J. (1982): *Spatial Processes in the Analysis of Vegetation*. PhD thesis, University of Western Ontario. National Library of Canada Microfiche No. TC 56124.
- Podani, J. (1984): Spatial processes in the analysis of vegetation. Theory and review. *Acta Bot. Hung.* **30**: 75-118.
- Scoggan, H. J. (1978): *The Flora of Canada. I-IV*. Nat. Mus. Nat. Sci. and Nat. Museums of Canada, Ottawa.
- Sneath, P. H. A., Sokal, R. R. (1973): *Numerical Taxonomy*. Freeman, San Francisco.
- Soó, R., Jávorka, S. (1951): *A Magyar Növényvilág Kézikönyve. I-II*. (Manual of the Hungarian Flora.) Akadémiai Kiadó, Budapest.
- Szőcs, Z. (1977): Újtípusú számítógép-orientált módszerek a vegetáció szerkezetének elemzésére (New computer oriented methods for the study of vegetation structure). Thesis, Vácrátót.

- Szűcs, Z. (1979): New computer-oriented methods for the study of natural and simulated vegetation structure. In: Orlóci, L., Rao, C. R., Stiteler, W. M. (eds): *Multivariate Methods in Ecological Work*. International Co-operative Publ. House, Fairland, Maryland, 301-308.
- Wensel, L. C. (1975): The treatment of boundary-line overlap in a forest sampling simulator. *Hilgardia* **43**: 143-159.
- Wensel, L. C., John, H. H. (1969): A statistical procedure for combining different types of sampling. *Forest Sci.* **15**: 307-317.
- 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.