

SPATIAL PROCESSES IN THE ANALYSIS OF VEGETATION: THEORY AND REVIEW

J. PODANI

RESEARCH INSTITUTE FOR BOTANY, HUNGARIAN ACADEMY OF SCIENCES, VÁCRÁTÓT, HUNGARY

(Received: 15 July 1983)

This paper is concerned with the time-static spatial dynamism in the analysis of vegetation. The main objective is to provide a theoretical framework for the discussion and future use of spatial processes. A process is defined as a sequence of changes in real or abstract space associated with a vegetation survey. The different spaces also constitute a series, permitting the distinction between two basic types of processes. The primary processes are defined by the investigator. Once a primary process is defined, dependent processes are automatically generated in the subsequent spaces. Further distinction is made according to the number of controlling parameters simultaneously changed during the process. The role of different processes is investigated in real, data, resemblance, classification and ordination spaces, and in the space of derived variables. Past contributions to the topic and the task of examining spatial dynamism are reviewed.

I. Introduction

The idea of process in vegetation has usually been tied with a sequence of spontaneous changes that occur in real space and time. During succession, for instance, the replacement of one species by another and the colonization of bare areas, as a result of the reproduction, propagation, growth and competition of species, represent the spatial-temporal dynamism. In the study of vegetation, however, the surveyor is often concerned with changes in real space that are not related to time. When the characteristics of sampling or some parameters of pattern analysis are subjected to successive modifications, time-static processes are defined. Although these processes are artificial, they do reflect inherent characteristics of the plant communities being studied.

It is important to realize that the term space is not necessarily limited to real three-dimensional space. Recently the use of derived variables (e.g., diversity) and multivariate techniques has become widespread in plant ecology, therefore abstract spaces are automatically involved even if the investigator is not aware of this fact. In abstract spaces different types of changes may be defined and used. Most of these are time-static, such as data transformation and standardization, or the successive modification of controlling parameters in the algorithms of multivariate techniques.

The analysis and discussion of time-static processes in real and abstract spaces appear to be of fundamental importance for three main reasons:

(1) The development of new numerical techniques has received increased attention in recent years. The use of multivariate methods is now a daily practice, but some classical problems of phytosociological sampling has become largely ignored. As Szócs (1979) pointed out:

... there is a striking contrast between the exactness of methods which analyze the data and the roughness of the procedures by which the vegetation is sampled. ... the quality of sampling has influence and limiting power on the information flow in the course of a complex vegetation study.

Plot size is perhaps the best example illustrating the general situation. About 30–40 years ago, and even up to the early 70's, the relationship of plot size to different variables and population parameters (e.g., number of species, variance/mean ratio) was a fairly popular research topic. For some reason, however, the total effect of sampling conditions on the results of multivariate analyses has been neglected except in some specific aspects (KERSHAW 1961, NOY-MEIR et al. 1970, FEKETE and Szócs 1974, and MATTHEWS 1978, 1979).

(2) The effect of sampling design seems even more important in succession studies which involve ordination and classification. Natural vegetation changes are to be shown and assessed, but these may only be revealed by an appropriate sampling design. Consequently, the adequacy of succession studies appears doubtful if spatial processes are not taken into account. Plot size is an example. The use of permanent quadrats with fixed size is typical in recent succession surveys (see e.g., FALINSKI 1977, VAN DER MAAREL 1980). A criticism of this practice is that the selected plot size will not always yield the maximum information to reveal successional changes (cf. JUHÁSZ-NAGY and PODANI 1983). The plot size optimal for describing the vegetation being studied may change as succession progresses. Therefore, simultaneous application of nested plots increasing in size may be appropriate. In this manner the effect of sampling at each point of time may be evaluated so that the investigator is enabled to recognize temporal changes.

(3) It is generally accepted that computer-oriented multivariate techniques are superior to the traditional methods because their results are less affected by subjective judgment. This seems, however, a self-deception, since the user of multivariate methods is also faced with arbitrary decisions. The choice of data type, transformation and standardization methods, resemblance coefficients, clustering and scaling algorithms may seriously influence the results. This effect is not always predictable even though the mathematical properties of most techniques are known. Despite the inherent objectivity of methods, the investigators' preference in favour of some particular technique will introduce much subjectivity into the survey. In this context, the evaluation of processes in abstract spaces is necessary to see how the results are affected by changes in the data and by slight modifications applied to the algorithms.

In this paper an attempt is made to establish a conceptual basis for the description and evaluation of spatial processes. First, the terminology of space, process, classification and ordination is discussed, and then the different types of processes are described. A review of the literature is embedded in the discussion in order to show existing approaches and to find their connections to the theoretical framework presented here.

2. Some basic terms

In the Introduction many technical terms were used assuming that their intuitive meanings were not doubtful. At this point, however, it is necessary to clarify the terminology concerned with spatial processes in other than the statistical sense. Terms, such as classification and ordination will also be discussed to find unambiguous definitions for later use.

2.1 Space

The Oxford Dictionary defines space as a "continuous extension viewed with or without the existence of objects within it" (FOWLER and FOWLER 1956). The most general concept of

space is topological (BAUM 1964). Since the different types of topological space are defined by axiomatic reference to relative positions of points, a space without objects is a nonsense. In a vegetation survey each methodological step is associated with a particular topological space (see WILDI and ORLÓCI 1980). The potential pathways of a study and the related spaces discussed in this section are summarized in Fig. 1. The first stage is of course concerned with the real space. The study area represents the three-dimensional physical space which may be simplified and transformed in some way to two-dimensional vegetation maps ("geometric images", Szócs 1979). The most critical step of the study, the sampling, may be performed on either the original or the mapped vegetation.

Abstract spaces are involved in all subsequent steps. By determining the state of each relevant variable (e.g., the number of individuals of species) for each sampling unit, the data (attribute) space is defined. Its dimensionality is potentially much higher than that of the real space. In data space the axes are variables (species) and sampling units are represented by points (species space, GITTINS 1969) or vice-versa in accordance with the attribute duality principle (LAMBERT and DALE 1964). The coordinates of points are the raw scores which are usually written in a species-by-sampling units data matrix.

Based on the use of appropriate objective functions the resemblance space is obtained in which the relative positions of points usually reflect pairwise dissimilarities, distances, etc., between the corresponding points of data space (resemblance structure, ORLÓCI 1978a). The resemblance structure of m objects is commonly expressed by an $m \times m$ matrix. If this matrix is subjected to linear decomposition, the maximal intrinsic dimensionality of resemblance space will turn out to be one less than the number of points or, if it is smaller, the number of axes in data space. A resemblance space is metric if all metric axioms (see ORLÓCI 1978a, p. 44) are satisfied. This is the case for many resemblance coefficients, but there are some exceptions. For example, two distinct points in data space are not necessarily distinguished by a chord distance (ORLÓCI 1967) in resemblance space (semi-metric space). Other functions may not fulfill the triangle inequality axiom (Sørensen index) or the symmetry axiom (Kulczynski index and I-divergence information) and the resemblance space will be non-metric.

If interest lies in meaningful properties of the data other than the resemblance structure of points, a new space will result which is the "space of derived variables". Its dimensionality equals the number of derived variables encountered. The term "derived variable" is used here in the broadest sense referring to any population parameters and sample descriptors (4.2) obtained by transformation. For example, the overall characterization of data in terms

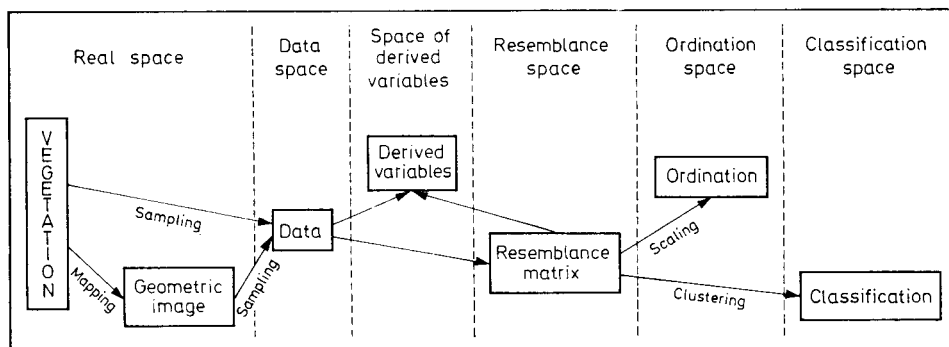


Fig. 1. Spaces and pathways in vegetation analysis

of diversity, sum of squares, expected resemblance, etc., leads to this space. The points represent data sets or subsets.

The different techniques of multivariate analysis summarize or disclose relationships among the objects in resemblance space. Scaling methods produce new coordinates transforming the objects into ordination space. If the ordination space preserves the resemblance structure, i.e., the relative positions of points remain unchanged, the dimensionality equals that of the resemblance space (e.g., principal components analysis, SEAL 1964). Some scaling methods destroy the resemblance structure in order to find a solution for an arbitrarily defined number of dimensions (e.g., multidimensional scaling, KRUSKAL 1964).

Cluster analysis produces the arrangement of points in classification space. When hierarchical relationships are established, the classification space has a maximum of $m-1$ dimensions and the distance of points satisfies the criteria for being ultrametric (ROHLF and SOKAL 1981) even if the resemblance space is non-metric. This is a strong argument against the use of non-metric coefficients in cluster analysis. In non-hierarchical classifications the between-group relationships are usually undefined and the intrinsic dimensionality of the classification space will be one less than the number of groups. As far as the presentation of results is concerned, the possible large dimensionality of classification space causes no problems: two dimensions are always satisfactory. Ordinations, however, cannot be shown in the plane properly since only two (or three, in case of stereograms, FEWSTER and ORLÓCI 1978) dimensions can be portrayed by a single diagram.

The discussion of conceptual spaces of plant ecology would be incomplete without mentioning the ecological space. Whereas the dimensions of ordination space are mathematical constructions, the axes of ecological space (WHITTAKER 1967) are defined as environmental gradients. Ecological space may be derived directly from the data when environmental gradients are known (gradient analysis, WHITTAKER 1967). In other situations ordinations may also lead to ecological space if ordination axes are identified as environmental gradients. Ecological space is considered as a desired output of ordination techniques (GAUCH 1982), therefore no explicit distinction will be made between ordination and ecological spaces in the sequel.

2.2 Process

The use of this term will be restricted to a series of changes strictly ordered by some internal or external criterion. A well-defined ordering is meant. A randomly or arbitrarily arranged sequence of changes is not considered a process. This restriction is consistent with the intuitive meaning of ecological processes which almost always suggest a spatio-temporal dynamism with interest focused upon events in time-domain (e.g., WATT 1947, SHUGART 1978, WHITE 1979, GRIME 1979). Succession, population growth and other changes monitored over time provide data on time series.

In spatial processes, events or changes associated with real or abstract spaces are central in importance, no matter whether time is involved or not. In statistical ecology, stochastic changes for generating spatial patterns are traditionally termed as spatial processes (BARTLETT 1975, CORMACK and ORD 1979, CLIFF and ORD 1981). For example, a completely random arrangement of points on a plane is considered to be the realization of a two-dimensional homogeneous Poisson point process (DIGGLE 1979) which develops by a series of random allocation of points in two-dimensional space. It is important that the term spatial process will be used here in a very different sense. Since pattern generating processes are not discussed here, the different terminology should not cause confusion. In the sequel, a spatial process will refer to any well defined series of spatial changes which is related to the analysis of vegetation data obtained at a given time. In this sense, the term spatial process was first used by JUHÁSZ-NAGY and his co-workers (JUHÁSZ-NAGY 1967, DÉVAI et al. 1971, JUHÁSZ-

NAGY et al. 1973). They examined association processes in space, i.e., the change of information statistical measures of interspecific relationships (see 4.2.9) as a function of sampling unit size. JUHÁSZ-NAGY's concept of spatial processes will be extended to cover many other sequences in real and abstract spaces.

Figure 1 illustrates different spaces encountered in a definite order in the course of a vegetation survey. Once a process is established in a given space, an unequivocal process will be automatically defined in subsequent spaces. When considering changes in sampling unit size, the series of plot sizes yields a series of data sets, resemblance matrices, classifications, etc. The *a priori* defined process, in the example of the increasing plot size, will be called a *primary process*. This is always artificial, the increments the sequence are selected according to design. The other processes defined automatically will be termed *dependent processes*. The change of interspecific association over plot size serves as example. It is to be noted that primary processes may also be defined in abstract spaces, for instance, specific data transformation functions applied to generate a process in data space. The primary and dependent processes will be discussed separately in sections 3 and 4.

2.3 Classification and ordination

In the general sense, classification is considered a process of allocating or ordering entities into initially undefined groups on the basis of their relationships (CORMACK 1971, EVERITT 1979, JARDINE and SIBSON 1971, ORLÓCI 1978a, SNEATH and SOKAL 1973). This should be distinguished from "identification" which is the way of assigning new, unidentified entities to previously established groups. As SNEATH and SOKAL (1973, p. 3) point out, there is some confusion over the meaning of classification in the literature:

Classification . . . is the name of a process; however, it has also been used for the end product of this process. Thus the result of classification is a classification.

In the preceding paragraphs classification was understood in the latter sense, primarily concerned with processes other than those involved in a classificatory algorithm (i.e., a defined sequence of computational steps). To avoid possible confusions, any process leading to classifications will be called cluster analysis or clustering, and the term classification itself will be restricted to the results.

The term "ordination" was first used by GOODALL (1954). The terminological problems are analogous with that of classification. To CORMACK (1971), ordination is a process of representing entities as points in a Euclidean space. SNEATH and SOKAL (1973) define ordination as the placement of entities (operational taxonomic units, in their terminology) in an abstract space of varying dimensionality. Ordination is often used by others to mean a procedure of ordering entities along axes of variation in a one or multidimensional scheme in such a manner that this arrangement will reveal some useful information about the relationships among the entities (GREIG-SMITH 1964, POOLE 1974, VAN DER MAAREL 1979a). Furthermore, the results of component analysis, factor analysis, reciprocal averaging, that is the scaling methods, have also been frequently referred to as ordinations. The later usage will be followed which is consistent with the convention accepted for classification.

3. Primary spatial processes

The primary processes are not natural, they are artificially generated to examine some intrinsic property of the vegetation being studied or to evaluate sampling and data analytical designs. The primary processes have simple

models related to questions raised by the investigator, such as: what is the effect of the data type upon ordination? Which quadrat size might be best used to detect aggregation of individuals of a given species?, and so forth. Most primary processes are confined to the real and data space. However, they are also conceivable in the subsequent spaces of a study, where interest usually centres on the dependent processes.

The primary processes are defined by specifying the controlling parameter or parameters and the sequence of their transformations. Two basic types can be distinguished based on the number of controlling parameters: In an elementary primary process, only a single parameter is modified. In a complex process two or more independent parameters are allowed to change. When a community is sampled by a set of circular plots, whose sizes increase, an elementary process is involved. The area of the plots is the controlling parameter. GREIG-SMITH's (1952, 1964) popular pattern analysis technique exemplifies a complex process since the size and shape of blocks are simultaneously changed and these factors are independent. In this case, there is a confounding of the effects of block size and shape on, for instance, the covariance of the two species being studied (see 4.2.11 for more information).

3.1 Primary processes in real space

In classification and ordination surveys, as well as in the analysis of the spatial pattern of plant species, the most commonly used technique for data collection is quadrat or plot sampling (it is noted that terms quadrat and plot will be used interchangeably). Quadrat sampling has four basic characteristics: size, shape, number and arrangement. The successive changes in any of these parameters represent an elementary process in real space.

Quadrat sampling, as PIELOU (1977) notes, suffers from the drawback that quadrats are not "natural", but are necessarily arbitrary units, and any of the four characteristics mentioned above may have considerable impact on the findings. Since there is no general substitute for quadrat sampling in Phytosociology, it will remain the principal and most direct technique to obtain vegetation data. This fact emphasizes the importance of the study of spatial processes.

Another technique for data collection is plotless sampling (COTTAM 1947, COTTAM and CURTIS 1949), but its phytosociological applicability is limited to specific problems (e.g., the study of interspecific relationships). In plotless sampling elementary processes may also be defined by changing some parameters of the design in nearest neighbour analysis (see 3.1.5).

3.1.1 Plot size

There is a general agreement between vegetation scientists that plot size has influence on the data and results. This may be the most substantial among the four plot characteristics (CAIN and CASTRO 1959, KERSHAW 1961, 1973, NOY-MEIR et al. 1970, ORLÓCI 1978a). EVANS (1952) concluded based on a pattern study that

... disagreement in the results of various authors stems in part from differences in the sizes of their quadrats, and success or failure [of their study] ... depends upon a fortunate or unfortunate choice of quadrat size.

Later GREIG-SMITH (1964) made similar comments demonstrating that the situation did not change much since EVANS' work:

The effect of sample size [i.e., sampling unit size] on indications of association has been largely overlooked. . . . the information derived from association data based on a single size of sample [plot] will be very incomplete and difficult to interpret.

The analysis of spatial pattern and the measurement of interspecific association are just examples, the question of adequate sampling unit size may be addressed in any other applications of sample plots in plant ecology. Estimations for population parameters (frequency, density, etc.) and their significance tests, derived variables (e.g., diversity), as well as ordinations and classifications should be treated with caution if based on a single plot size. The analysis of processes depending on the area of the sampling units may therefore contribute much to the understanding of fundamental theoretical and practical questions of sampling.

In Phytosociology, the question of adequate plot size is inevitably linked to the concept of "minimal" or "characteristic" area (cf. GOODALL 1970). The literature of minimal area is voluminous and somewhat confusing. TÜXEN (1970) published an extensive bibliography of relevant articles up to that time. It seems that minimal area could have been the sole topic of another paper. Here, only a few relevant points are made.

Almost all definitions agree in that the "true" characteristics of a community first appear at the size of minimal area and its increase does not yield significant amounts of additional information (GREIG-SMITH 1964). In general, minimal area is thought to be the smallest plot size representing the "characteristic structure" and composition of the plant community (CAIN and CASTRO 1959, GOODALL 1961). Differences arise in the selection of characteristics that are considered to be important in determining minimal area. WESTHOFF and VAN DER MAAREL (1973) pointed out that a sharp distinction has to be made between minimal area as a synthetic concept and the analytical minimal area. The synthetic minimal area [SMA] is defined in many ways according to the principles of different phytosociological schools, reflecting various, sometimes contradictory views. For example, the Uppsala school of DU RIETZ (1921) relied on a circular definition using constancy:

The minimal area is that in which all the constants occur, and the constants occur in the minimal area. . . . the constants characteristic of each association [i.e., plant community type] are already noticeable in case of relatively small quadrats, while in further enlarging these the constants do not increase in number.

(TRASS and MALMER 1973). WESTHOFF (1951) gives another definition for SMA:

. . . the minimal surface which has as a rule to be occupied by a sample of a plant community if the normal specific assemblage will be able to develop.

Without continuing with citations, it is clear that the concept of SMA is meaningful only *after* the associations or types have been established. Furthermore, it is almost always implicitly assumed that sample plots of the size of minimal area, originating from different stands assigned to the same abstract syntaxonomical unit, are relatively very similar to one another. In other words, the set of such plots has to show maximum homogeneity (in the sense of overall similarity). Consequently, if these plots are laid down within a given stand of a plant community, the sample is expected to be even more homogeneous. However, as JUHÁSZ-NAGY (1967) pointed out, it is more useful to detect an area at which the stand is found in its richest and most diverse development than to look for SMA. This area is called maximum area [MA] and JUHÁSZ-NAGY suggested information theoretical functions for its determination (see 4.2.9). MA is analogous to SMA in that it is also an *a posteriori* concept, it may be

defined for a stand which, as a complete entity, represents an already defined plant community type. For this reason, the collective term *characteristic area* [CA] may best be used.

The *analytical minimal area* is generally regarded as the "optimal" size of sampling units by which the vegetation should be sampled for subsequent analysis. Since this refers only to a minimal area, it seems more appropriate to use the term "*analytical characteristic area*" [ACA]. A fundamental difference between CA and ACA is that when ACA is to be determined, the vegetation types are not yet known. It is obvious that the correct recognition of types, if they exist at all, may be achieved only if the individual CAs are best approximated by ACA. Also, the sample has to show minimum overall "*homogeneity*", otherwise the differences among the possible types are not completely manifested in the data. Therefore, ACA is consistent with MA and inconsistent with SMA, indicating that MA is perhaps the more useful of these concepts in typification.

In an optimal situation, which is rather hypothetical, the CAs of the types to be recognized are equal, therefore if ACA is found, the CAs are also found and the classification is not in doubt as far as plot size is concerned. In practice, however, CA potentially differs with community types and any attempt to detect the types necessarily involves a circular argument: optimum plot size for each type can be determined only if the types are already known, but the adequate recognition of the types is conditional upon the use of sampling units of the optimum size. Moreover, if vegetation types are arbitrarily established and the CAs are determined in some way, the existence of classes could not be justified by multivariate analysis without violating the statistical assumption that plots of different size are not subjected to simultaneous scaling or cluster analysis. For these reasons, the only theoretically justifiable strategy to indicate types on the basis of their individual CAs is an iterative approach for which an example is given in 4.4.1. Fortunately, when the classifications appear less influenced by plot size, there exists a simpler alternative which does not require the determination of characteristic areas. It is discussed together with the dependent processes of the classification space (4.4.1).

Many different techniques have been suggested for pattern detection or to determine characteristic areas. All of them rest on the analysis of processes dependent on the area of plots. There are two different possibilities to modify the size of sampling units keeping in mind that the other three characteristics of sampling are held constant. These are:

1. *Nested plots.* Each plot is successively enlarged such that the largest plot will completely include $k-1$, the second largest $k-2$ plots, and so on, if k is the number of different sizes (Figs 2a-d). In this case the data and results obtained at different plot sizes are not independent from one another. This may be a disadvantage, since statistical tests for comparing results may not be valid, although the examination of dependent processes may still give useful information (e.g., cumulative species/area curves). It is not to be forgotten, however, that the nesting of plots is a prerequisite to the assessment of the effect of plot size upon, for instance, resemblance structures.

2. *Independent plots.* Each set of sample plots of a given size is laid down separately and randomly to ensure independence between plot sizes. The results can be tested for significance but the price is high: the sampling requires much more effort than the nested plots. Further difficulty is that the

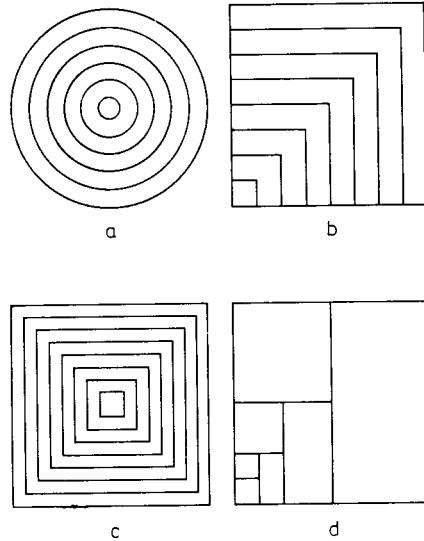


Fig. 2. Different designs for nested plots

vegetation is more disturbed or may be completely destroyed due to intensive trampling.

The smallest and largest plot size, as well as the increments are arbitrary. The choice is affected by economic considerations and one's knowledge of the vegetation to be studied.

3.1.2 Plot shape

The shape of a sampling unit may be a cause for fewer problems than plot size. Although there are some reports treating the effect of plot shape (e.g., CLAPHAM 1932, MYERS and CHAPMAN 1953, VAN DYNE et al. 1963), this problem was largely ignored. In Europe, isodiametric units have long been used. The shape is usually square since it has the great practical advantage in that the marking of plots is easily done in the field. Another reason for the preference for square-shaped units is that they can be arranged into a continuous grid (commonly used in pattern studies). Circular plots, although just as natural, are used on rare occasion in actual studies (e.g., JUHÁSZ-NAGY 1967). In North America short rectangles of $20 \times 50 \text{ m}^2$ size (the so-called tenth-hectare plots), subdivided into smaller units are frequent, especially in gradient studies (cf. WHITTAKER 1973a).

In a heterogeneous vegetation the optimum plot shape depends on the objective of the survey. For classification and ordination purposes isodiametric plots are recommended; because the probability that a plot will overlap community-type boundaries (or other vegetation discontinuities) is the smallest. Therefore, it is expected that the within-plot variance (variability, heterogeneity) is kept to the minimum, whereas the between-plot variance is maximized. If precise estimates for population parameters have to be derived from the whole sample, the opposite condition is aimed at, i.e., maximum within-plot and minimum between-plot variances. This may be achieved by the use of elongated sampling units.

A spatial process with respect to shape may be the successive elongation of square plots to long rectangles or of circular plots to long ellipsoids of the same area. The elongation of shape, however, increases the boundary/area ratio thereby introducing undesirable edge

effects (GREIG-SMITH 1964). This fact has to be considered when analyzing processes depending on the elongation of sample plots.

3.1.3 *Sample size*

The number of sampling units is the size of the sample. This is critical in surveys where the objective is estimation of vegetation parameters and statistical inference. The robustness of most statistical analyses against the violation of assumptions largely depends on sample size. A few plots often yield high standard error of the mean and, even if the statistical tests are significant at conventional probability levels, the confidence limits may be undesirably wide. GREIG-SMITH (1964) presents an example. The most efficient way to increase precision is to increase the number of plots, although the other three characteristics of sampling may also be changed to achieve this goal (see 4.2.2). This procedure generates an elementary process in real space. In plotless sampling the analogous process is obviously the enlargement of the sampled area.

The number of plots is closely related to sampling effort, so a balance between the precision of results and the costs involved is critical. This problem requires the study of processes depending on sample size. A starting point may be the common observation that successive equal increases in the number of plots do not yield proportional gains in the additional information. The point where the new information (or the change of results) becomes negligible may be determined by specifying the level of the desired accuracy and performing a pilot study. Of course, sample size is not just a practical question. If, for instance, the relationship between the number of plots and a derived variable is not known, the sample size has to be changed until the nature of their relationship is safely recognized.

The problem of sample size has to be treated separately in ordination and classification when the sampling units themselves are subjected to analysis. In these cases the goal is not estimation and significance testing but typification or trend seeking. There is, however, no study and consequently no evidence suggesting that the increase of sample size improves the final results. Too few plots are obviously unsatisfactory because the data will not contain sufficient information for the analysis. On the other hand, a very large number of plots will probably increase random variation (noise) which may significantly affect the results. Also, a too large data set may be difficult and too expensive to analyze even at the present advanced state of computer technology. The optimization of sample size prior to scaling and clustering therefore requires further consideration.

The influence of sample size on the stability of results obtained by multivariate techniques is to be also mentioned. One aspect of stability is the robustness of ordinations and classifications (ROHLF and SOKAL 1981) which is satisfied if the addition of new objects to the sample does not alter significantly the groups and/or trends. The robustness of phytosociological classifications and ordinations may be examined by successively increasing the number of plots analyzed and examining the resulting changes. Whereas the numerical taxonomic analogue of this problem is a central issue in systematics (cf. CROVELLO 1968), vegetation scientists largely ignore stability/sample size relationships. Although GOFF (1975) investigated the effect of the inclusion of new species on a reference ordination of other species, it seems that there are no publications in the literature concerned with the modification of sample size in multivariate analysis of vegetational plots.

3.1.4 *Spatial arrangement of sampling units*

The placing of plots in the field is subject to the goal of the survey and the personal experience and preference of the surveyor. In plant ecology, four different sampling designs are commonly used, the preferential, random, systematic and restricted random. Other

methods, e.g., cluster sampling, are seldom used. Before discussing the possible processes related to the arrangement of plots, it is worthwhile to summarize the main features of each sampling technique.

The traditional sampling method used frequently and mostly by European analysts is preferential. After a reconnaissance of the survey area those sites, which appear reflecting the most characteristic structure of the community, are chosen and sampled. However, as GREIG-SMITH (1964, p. 21) pointed out:

Selection of typical sites for samples is clearly inappropriate to a quantitative approach, as their choice is dependent on the observer's preconceived ideas of the character of the vegetation, and data from such samples cannot be considered an unbiased estimate of the vegetation of the area.

Although preferential sampling is not necessarily bad, no processes can be logically defined in this method.

In simple random sampling any part of the study area has the same probability of being sampled and the placement of a plot is completely independent of the positions of those already located. The randomized arrangement ensures that the sampling is uninfluenced by arbitrary decisions; the data obtained in this way represent the most powerful input for statistical analyses. Randomness may be easily achieved by computerized sampling based on mapped and digitized data. However, the random sampling technique is less straightforward in the field if the sampled area is large.

Whereas randomization is considered the best strategy to obtain measurements for overall properties of a homogeneous vegetation, systematic sampling may provide more information on variability within an area (GREIG-SMITH 1964). In systematic sampling the plots are arranged along a regular network of points (lattice). Of the many variants of systematic sampling (see SAMPFORD 1962) the special case in which plots are placed at all points in a square lattice seems important and most widely used in vegetation studies. Term systematic sampling will be used to describe this special case. Sample size, area and inter-plot distance (spacing) are not independent in systematic sampling; and two determine the third. In many applications the plots are contiguous and they completely cover a small study area (grid). A special case is the belt-transect commonly used in pattern studies. An obvious disadvantage of this and other systematic sampling is that the position of the first lattice point determines the others, and if the whole lattice is shifted, the results obtained may be greatly different (see 4.2.11).

As an intermediate sampling method between the random and systematic methods, the use of restricted (stratified) random sampling is suggested by many authors (see SAMPFORD 1962, GREIG-SMITH 1964, GREEN 1979). This method combines the advantages of the other two sampling designs in that the study area is divided into large compartments (strata, blocks) within which the plots are placed at random. The number of plots within each block is usually proportional to its size (proportional allocation). GREIG-SMITH points out that the most advantageous variant of this strategy is obtained if the blocks are equal and relatively small in size and only one plot is laid down within each. If the number of plots to be placed within each strata is determined in such a way that the more variable blocks are sampled more intensively, a considerable increase in precision may be achieved (optimal allocation). This technique requires the knowledge of variances for each stratum. This can be determined through preliminary sampling.

It will be shown below that the objective sampling methods are in fact different stages of a series which might be termed sampling process. The process starts with the random location of a point within the study area and the determination of the other points of a lattice. In the first stage the sampling units are placed at these points as centroids (Fig. 3a) so at

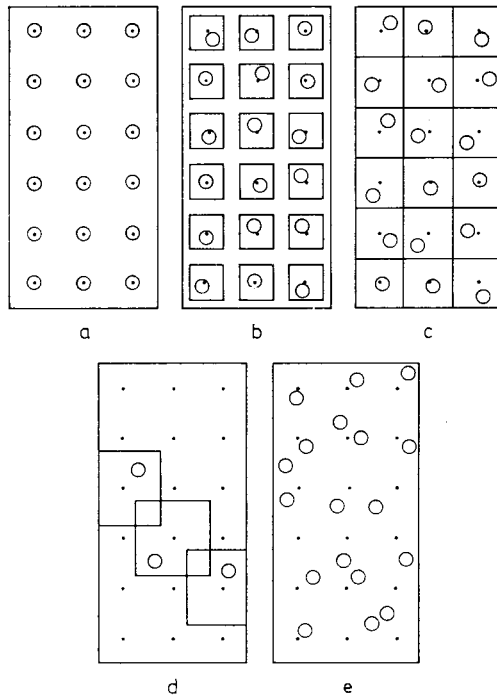


Fig. 3. Illustration of the sampling process with square compartments and circular plots. *a:* Systematic sampling. *b:* Intermediate stage. *c:* Stratified random sampling. *d:* Intermediate stage (only three strata are shown). *e:* Random sampling

this stage the sampling is systematic. In the subsequent steps the lattice points will serve as centroids of square compartments increasing in size. Within each compartment one sampling unit will be laid down randomly (Fig. 3b). When the compartments become contiguous, the design is equivalent to the most powerful variant of the restricted random sampling (Fig. 3c). The further increase of blocks results in overlapping squares, therefore the arrangement of plots will approximate more and more the randomized configuration (Fig. 3d). When all blocks become large enough to include the sampled area, the sampling units will be completely random (Fig. 3e). Obviously, the execution of this process in the field would be impractical if not impossible. Computer based sampling techniques seem to be the most efficient means for the study of the continuous series of different plot arrangements. A FORTRAN program to perform such a sampling process using digitized field data has been developed by PODANI (1982).

Another possibility for a spatial process with respect to the arrangement of sampling units is the successive increase of spacings between replicate plots (GOODALL 1961) or between cells picked up from a grid or a belt-transect (GOODALL 1974, LUDWIG and GOODALL 1978). This elementary process may be used for the detection of aggregation of individuals and possible periodicities in spatial pattern (4.2.11).

Finally, the orientation of sample plots should be also mentioned as a potential controlling parameter for a process. If anisodiametric units (strips, ellipses) are employed for sampling, and the vegetation is heterogeneous in a particular direction, the orientation of plots will obviously influence the observations. Parallel orientation with the gradient is

expected to yield more precise estimates for the vegetational variables than plots perpendicular to the gradient, since the between-plots variance is higher in the latter case (cf. LAFRANCE 1972). For this reason, the rotation of anisodiametric plots may serve as a primary process to find possible uni-directional heterogeneity within the study area. Of course, this possibility is mostly theoretical since scaling methods are more efficient in gradient analysis.

3.1.5 Nearest neighbors

The difficulties with the choice of an optimal sampling design when plots are the units, have encouraged plant ecologists to suggest the use of plotless sampling in vegetation surveys. Of the many variants of this technique (see GREIG-SMITH 1964, PIELOU 1977, for details) the method of nearest neighbors offers a possibility to define primary processes. In the simplest case the distances from randomly selected plants or random points to the closest plants are measured and analyzed. CLARK and EVANS (1954), MORISITA (1954) and THOMPSON (1956) pointed out that the method which involves measurement of the distances to the second, third, . . . , m -th nearest plant is preferred (multiple nearest neighbors). Obviously, the successive increase of m represents a primary process in real space.

WILLIAMS et al. (1969) suggested a version of multiple nearest neighbor analysis, applicable to the study of multi-species populations. The exact positions of all n individuals on the sampling area are determined and all inter-plant distances are calculated. Each plant, together with its m nearest neighbors, is then regarded as a sampling unit. The set of the n overlapping units (point-clumps) can be subjected to multivariate analysis. It was found by WILLIAMS and his co-workers that the information content of a point-clump system is entirely a function of the mean distance from the reference individuals. This fact suggests that a hybrid sampling procedure may be also appropriate in pattern studies. Each individual may be considered the centroid of a circular plot with radius r . All plants falling within this plot are then recorded, irrespective of quantity. In this sampling strategy the change of r constitutes a process.

3.1.6 Complex processes

When a simultaneous modification of two or more criteria is allowed in sampling, one has a complex primary process in real space. Clearly, the number of possibilities is large, but only one case seems important. This process is the series of block sizes implied in grid-analysis originally suggested by GREIG-SMITH (1952) and widely applied to pattern studies (see GREIG-SMITH 1964, 1979, KERSHAW 1973, and references therein). The starting situation is a grid of $2^s = 2^k \cdot 2^{k+1}$ square units, where s and k are integers. The number of individuals, cover, or some other variable is recorded in the units. Adjacent pairs of cells are then combined to obtain 2^{s-1} oblong blocks of two units; adjacent two-unit blocks are next merged to give square, four-unit blocks, and so on. That is, whereas the blocks are doubled in size at each stage of this process, they are continually alternating in shape from a square to a short rectangle of proportion 1 : 2 (Fig. 2d). The consequences of this complexity will be discussed in 4.2.11. If the successive combinations of adjacent units lies within a belt transect (KERSHAW 1957, USHER 1969, HILL 1973a) the process is also complex with a constant block size/block length ratio. A modification of the original pattern analysis technique (FEOLI et al. 1982) may lead to an even more complex series since it allows the blocks not to be uniform in size and shape at any given stage of the process.

3.2 Primary processes in data space

The raw vegetation data may be directly subjected to numerical analysis without any change, but in most cases, simplifications and/or transformations are advantageous for dif-

ferent purposes. Data simplifications can reduce sampling effort and the costs of the survey, whereas the rationale for transformations is rather theoretical. Manipulations of the data, if the changes can be arranged into a well-defined series, are capable of generating primary processes. Four categories of these processes will be distinguished: reduction of the dimensionality of data space; changing the number of points in data space; simplification of data type; and data transformation.

3.2.1 Reduction of dimensionality

Many multivariate methods cannot be programmed and the computations cannot be performed without ensuring an easy and continuous access to the raw data during the calculations. Since the storage capacity of the computer available to the user determines maximum problem size, the analysis of some data sets may be impossible. Even if the maximum is not exceeded, some programs are very expensive to run if the number of variables is high. In some cases the computational difficulties may be diminished by reducing the dimensionality of data space. An efficient reduction may be accomplished by principal components analysis which eventually leads to an ordination space, since the original variables are replaced by independent components. However, instead of this indirect dimension-reducing procedure, relatively simple techniques which enable the user to recognize and delete the least important variables will be discussed.

It is a common observation that plant communities may be satisfactorily described based on a considerably reduced number of species. For example, AUSTIN and GREIG-SMITH (1968) concluded that ordination of rain forest communities using less than 25% of the total flora was efficient and interpretable. WEBB et al. (1967) found that a subset of large trees duplicated the original classification of sites perfectly, and no other subset possessed this property. The reason given is that the minimum information sufficient to obtain meaningful results may be provided by relatively few species, the others are redundant (KAESLER et al. 1974) and, which may be even worse, too many species may contain much noise potentially obscuring the clarity of ordinations and classifications (cf. JANCEY 1980). The subset of the best p species out of n could be chosen by examining all combinations but this method is clearly impractical in most cases (ORLÓCI 1973). A more suitable approach to this problem is the measurement of species importance (i.e., the determination of weights), a subsequent species ranking based on these values and the deletion of the less important species. The weighting and ranking procedures recently available have been reviewed by ORLÓCI (1978a), DALE and WILLIAMS (1978) and JANCEY (1979).

Species weights may be determined from the raw data only or from the data and results (*a priori* and *a posteriori* weighting, respectively). Clearly, the *a priori* weighting is now of principal concern, although *a posteriori* weighting may be also useful in an iterative approach to reduce dimensionality (and to diminish random noise, JANCEY 1980). A further categorization of weighting and ranking techniques can be made on the basis of their computational feasibility and simplicity. Some techniques require only simple examination of the data. The weight for each species is determined based on the summation of similarity values to the other $n-1$ species (e.g., MACNAUGHTON-SMITH et al. 1964), by summing the deviations from the expectations in the data matrix (DALE and WILLIAMS 1978) or by calculating the contribution of species to the total sum of squares, mutual information or other overall property of the data. Most of the weights obtained in this way will not be independent and, as a consequence, the highly correlated species will have similar ranks. This is not so with the more sophisticated ranking techniques suggested by ORLÓCI (1973, 1976a, 1978b), which may require $n-2$ algorithmic cycles to find the rank order. In the first cycle, in fact, a simple ranking is performed, the species with rank 1 is found and then removed from the data. The

second most important species will be selected in such a manner that it will be the most independent of the first, the third species will be the most independent from the first two, and so on. A rationale of this strategy is to find the minimum subset of species which accounts for the maximum variability in the data in terms of independent components. The analogy to principal component analysis is obvious, although in this case the correlations between variables are not eliminated but minimized. It is very important to note, that the ranking procedure used should be consistent with the subsequent operations. There is no point to analyze a data subset by sum of squares clustering if the species are ranked based on information theoretical criteria. However, the choice still remains subjective because the concept of sum of squares may be meaningful for several ranking methods. Therefore, an evaluation of the different strategies would be necessary by comparing their efficiency with respect to data reduction before clustering and scaling.

Having ranked the species, an arbitrary cut-off point can be specified and the less important species can be deleted. In this way a remarkable reduction of dimensionality may be achieved and the loss of information is minimized. Of course, besides ranking methods other techniques may be used to recognize unimportant variables. For example, species may be simply deleted if their cover estimate or frequency is less than a fixed value. In any case, the spatial process will be defined by changing the deletion criterion. The effect of species deletion at different thresholds upon the results can be evaluated and the point where the changes become significant can be determined.

An alternative way to reduce the number of variables is the fusion of dimensions. Phytosociological data are traditionally recorded at species level but several authors (VAN DER MAAREL 1972, DALE and CLIFFORD 1976, DALE 1977, DEL MORAL and DENTON 1977) suggest that the numerical analysis of vegetation data may be appropriate based on higher taxonomic categories. Data of species belonging to the same genus may be added or averaged, data of genera falling into the same family may also be combined in some way, and so forth. Such successive replacement of several dimensions with single ones will be a primary process in data space. PRELOU (1969) proposed a method to calculate the information content attributable to each taxonomic level. Data from preliminary sampling can be subjected to this analysis and the level where the information gain (or loss) is negligible can be determined for use in the subsequent survey.

Arguments in favour of coarse taxonomic ranks have primarily come from environmental impact studies. GREEN (1979, p. 118) gives three reasons for potential use of higher ranks. These are:

- (1) Species may prove to be mixtures of several sibling species (GRASSLE and GRASSLE, 1976, provide a zoological example).
- (2) The taxonomy is often poorly worked out at the species level therefore the identification at the generic level is easier and more reliable.
- (3) Genera may reflect major environmental differences whereas species may be specific to the micro-environmental patchiness in a site.

The use of higher taxonomic ranks is recommended by students of environmental pollution assuming that the response of related taxa to environmental differences is similar. However, this condition is not generally satisfied. Another difficulty is that taxonomic categories are subjectively created, and the hierarchical classification of plants may be considerably different depending on the personal judgement of the taxonomist. Criteria for creating genera and higher categories may vary even within the same classification. Finally, fusion of taxonomic categories implies an underweighting of species-rich taxa which lacks theoretical justification. For these reasons, the reduction of taxonomic ranks cannot be generally recommended in Phytosociology, even though there are some successful applications in large-scale surveys (e.g., DEL MORAL and DENTON 1977).

3.2.2 Changing the number of points in data space

On first thought the proposition of changing the number of points in data space might seem to be nonsensical, since results of multivariate analyses will most likely be different and difficult to compare. However, it was already mentioned in 3.1.3 that the change of sample size may yield very useful information about the robustness of multivariate analysis. In fact, the increase of sample size can also be considered as a process defined in data space. It is a common practice to subject the species to scaling and clustering, therefore the relevance of the stability of results is not restricted to the analysis of plots. GOFF (1975) provides an example stressing that the objective of phytosociological surveys may be to establish a reference ordination of given species from a diverse region. If it is so, then, it is an important point to determine the effect of additional species upon the configuration in the initial ordination. The increase of the number of species involved represents a primary process in data space.

3.2.3 Simplification of data type

The type of data collected in the field is critical in determining the sampling effort. In classification and ordination surveys it is therefore often advisable to find a simplified data type without much loss of information. This goal may be achieved in a pilot sampling by recording the most detailed quantitative data (e.g., counts, cover). After this the effect of a series of data simplifications on the results can be evaluated, and a less costly method of data collection for which the loss of information is negligible can be selected.

When the number of individuals is recorded, the definition of an upper limit of density, k , beyond which the individuals of each species are not counted may be used to reduce sampling effort. In this case we assume that the resemblance structure of sampling units or species will not be distorted if the counting stops at k . In the pilot study the plant populations within some sample plots have to be fully censused. The amount of distortion at different values of k can be estimated after truncating the counts such that each score larger than k is set equal to k (ORLÓCI and MUKKATTU 1973).

Cover data have to be estimated as precisely as possible in the pilot study. Afterwards, the scores can be converted to categorical data in steps according to scales of diminished accuracy. Each cover datum is replaced by the mean of the cover class within which the actual value has fallen. Another possibility is the substitution for cover estimates the ranks of the corresponding cover categories. In this case, an ordinal scale will result if the categories are unequal. Such scales are commonly used in Phytosociology. The ultimate data simplification will be a presence/absence (signum) transformation for all strategies. It is to be noted that the data simplification processes imply the equalization (see 3.2.4) of variables, since the differences between high and low values of species are gradually diminished.

If one desires to analyze presence/absence, and quantitative data are available, the variables must be dichotomized. Besides the self-evident way of dichotomization in signum transformation,

$$x' = 1, \text{ if } x > 0, \text{ otherwise } x' = 0 \quad (1)$$

one can also define the dichotomy 1/0 using non-zero thresholds as suggested by NOY-MEIR et al. (1970). According to their definition a species will receive score 1 in a sampling if its quantity exceeds a given threshold. As the authors note, "it would be desirable . . . to develop a procedure which would search for an optimal threshold for each species". Such a method might help the phytosociologist not only to reduce sampling effort but also to improve the results of multivariate analyses by diminishing "noise" in the data. However, the exact relationship between noise and presence thresholds deserves further research.

3.2.4 Data transformation

Transformation of data prior to multivariate analysis is inevitable if the variables are not commensurable, i.e., they are measured on different scales with different units, such as biomass and density. Similar manipulations may be necessary if the scale is common but there are undesirably excessive differences in the range of variables, for example, species with consistently low or high cover scores. In geometric terms, data transformation is the relocation of the points in data space by means of modifying the scale on the axes. The main types of transformation are equalization and weighting (ORLÓCI 1978a).

Through equalization differences between variables are diminished by adjusting the data to zero mean and unit variance, to unit range, etc. A widely employed method of equalization is the derivation of a new range for each variable which is intermediate between the original range and unity. Logarithmic, square-root or other monotone functions may be used for the latter purpose. Weighting is the assignment of a specified weight to each variable whereby its importance is changed. The weights may be determined arbitrarily or derived from the data. The terms equalization and weighting are inseparable in the sense that equalization implies overweighting of certain variables and underweighting the others whereas weighting may imply equalization of variables which originally were not equally important.

A primary process in data space can readily be defined by the proper selection of transformation functions of the general form:

$$x' = f(x, c) \quad (2)$$

where x is the original score, x' is the new value and c is a parameter selected by the investigator. Only a few of the many possible options for (2) will be mentioned. Logarithmic and power functions, in which c is an exponential quantity, may be applied to under- or overweight dominant species to the extent specified by c . An exponential function was proposed by R. CLYMO (in VAN DER MAAREL 1979a) for the transformation of relative cover data:

$$x' = (1 - \exp(-cx)) / (1 - \exp(-c)) \quad (3)$$

where $0 \leq x \leq 1$. For large values of c the presence/absence situation is approximated. When c is very small, x' is approximately equal to x . For large negative values of c species with high cover values are extremely overemphasized while the others are almost entirely ignored (Fig. 4). Note that this function has a singularity at $c = 0$.

3.3 Primary processes in resemblance space

It seems on first sight that a process in resemblance space could be defined in like manner as the transformation processes discussed in the previous section. That is, the resemblance function measuring distance, similarity or other relationship between objects i and j may have the form:

$$d_{ij} = f(X_i, X_j, c) \quad (4)$$

where X_i and X_j are data vectors pertaining to the respective objects and c is a parameter (or function) controlling the process. However, there appears to exist no published information, as to this author's knowledge, which treats primary processes in resemblance space. The reason may be that most resemblance functions used in ecology are not suitable to define a process. MINKOWSKI's generalized metric and RÉNYI's (1961) information of order c are obvious exceptions. The ecological meaning of the MINKOWSKI functions is not clear, however, except in two well-known forms (Manhattan metric and Euclidean distance). RÉNYI's information function can be used to generate a family of divergence measures.

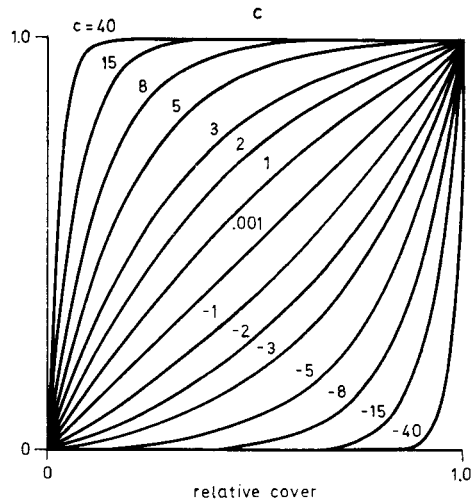


Fig. 4. Data transformation process using CLYMO's function

3.4 Primary processes in the space of derived variables

A process with respect to a derived variable, v , can be established using a function analogous to (2) and (4) given by

$$v = f(X, c) \quad (5)$$

where X is the data matrix and c represents the controlling parameter. The possibility of these processes is rather theoretical, most of the functions for derived variables are not applicable. One exception is RÉNYI's (1961) entropy of order c which might be used to define a primary diversity process. The ecological significance of changing the value of c has been studied (HILL 1973b).

3.5 Primary processes in classification and ordination space

A primary elementary process in classification and ordination space might be generated by successive small modifications of the algorithm. Of course, a number of methods are not suitable for this, since their algorithms are completely and unequivocally defined (as the majority of hierarchical clustering and the scaling methods). By contrast, in several methods one or more parameters need to be pre-determined by the user, so it is possible to set the control parameters equal to different values. All non-hierarchical clustering algorithms (JANCEY 1966, 1974, WILLIAMS and DALE 1965) and some scaling strategies (ORLÓCI 1980) are of this type. It is obvious that a great deal of arbitrariness is involved in the selection of parameters. This supports the view that the performance of these methods can only be evaluated by studying the influence of the series of algorithmic changes upon the results.

3.5.1 Hierarchical classifications

LANCE and WILLIAMS (1967) gave a common combinatorial formula to describe eight hierarchical clustering methods, defined by

$$d_{ij,k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}| \quad (6)$$

where $d_{ij,k}$ is the dissimilarity between group k and the group obtained with the fusion of groups i and j . d_{ij} , d_{ik} and d_{jk} are the corresponding pairwise dissimilarities (often squared Euclidean distances). The values of α , β and γ are strictly determined for seven strategies. In the eighth, the flexible sorting method, the user is faced with an arbitrary choice, and the parameters may be set to any value within the following limitations:

$$\alpha_i + \alpha_j + \beta = 1; \quad \alpha_i = \alpha_j; \quad \beta < 1; \quad \gamma = 0.$$

As β approaches 1, the analysis produces strong chaining similarly to the nearest neighbor methods. For negative values of β several well-separated and rather homogeneous clusters are obtained resembling the results of furthest neighbor sorting. There is no doubt that a series of hierarchies obtained with intermediate values is appropriate to reveal structural properties of the data.

3.5.2 Partitions

The results of non-hierarchical clustering are usually greatly affected by the selection of control parameters. In the simplest cases only a single parameter must be specified, e.g., a reallocation threshold after divisive clustering (CRAWFORD and WISHART 1968) or the number of clusters (FORGY 1965, JANCEY 1966, MACQUEEN 1967). In the latter case the dependence of results on the pre-determined parameter is obvious. In graph theory clustering, the user must make arbitrary decisions regarding the neighborhood radius by which simply or completely connected subgraphs will be obtained (ESTABROOK 1966, VAN GROENEVOUD and IHM 1974). A series of identical partitions may also be generated by intersecting single and complete linkage dendrograms at the corresponding hierarchical levels. Since any dendrogram represents a hierarchic system of partitions, all of them can be considered in a sense as graphical illustrations of non-hierarchical classification processes.

Some more sophisticated techniques (e.g., JANCEY 1974, ORLÓCI 1976b) are closely related to nearest neighbour analysis but they require the pre-determination of two or more parameters. These methods are applicable to test some *a priori* hypotheses about the existence of distinct groups (ORLÓCI 1978a). The rapid techniques aiming at the initial clustering of large data sets (JANSSEN 1975, GAUCH 1980) should be also mentioned. If only one control parameter is altered and the others are fixed, one can readily define elementary processes.

It is noteworthy that some non-hierarchical techniques do not necessarily produce unique results even if the same control parameters are applied. This is the case when the initial partition is generated arbitrarily or at random.

3.5.3 Ordinations

A primary process in ordination space can only be generated if the method requires the pre-determination of some control parameters. A good example is the procedure of parametric mapping developed by SHEPARD and CARROLL (1966). The method seeks a minimal number of dimensions (p) in which the points (representing species or relevés) are related by some response curves or surfaces which are as smooth as possible. For each predetermined value of p the VON NEUMANN quantity

$$K = \sum_i \sum_j \frac{(d_{ij})^\alpha}{(D_{ij})^\beta} / \left\{ \sum_i \sum_j (D_{ij})^\gamma \right\}^{-\beta/\gamma} \quad (7)$$

should be minimized, where d_{ij} and D_{ij} are the Euclidean distances between points i and j in the data and the ordination space, respectively. The parameters can have any values provided that $\gamma = \alpha - \beta$. Thus the solution depends on the specified number of dimensions and the arbitrarily determined values of α and β . As NOY-MEIR (1974) pointed out, the parameter β has a remarkable effect on the resulting ordination. The larger its value, the stronger is the weighting in favour of closely located points, therefore a local smoothness of response surfaces will be achieved. If β is rather small, a global optimum may be obtained.

Other ordination methods which should be mentioned in this context are the multi-dimensional scaling method proposed by KRUSKAL (1964) and a multidimensional curve-seeking procedure suggested by SNEATH (1966). Both methods have several parameters to be defined arbitrarily.

4. Dependent processes

The literature of the study of dependent processes is scattered, diverse and frequently unaccessible. Therefore any attempt to provide a complete review or just a bibliography would probably end in failure. Some specific questions have already received much attention by ecologists. For example, species/area and variance/block size relationships are very often examined regardless of the main objective of the different surveys. On the other hand, the study of dependent processes in resemblance, classification and ordination space seems less exhaustive, and there is more chance to cover all relevant work. For this reason, the following discussion will attempt to give a more balanced treatment of the topic.

The categorization of dependent processes relies on the nature of two spaces. Possibilities are shown in Table 1. Dependent processes do not exist in real space and they are of little importance in data space; their study is almost completely confined to the others.

Table 1
Possible combinations of primary and dependent processes

Space of primary processes \ Space of dependent processes	Space of dependent processes					
	Data	Derived v.	Resemblance	Classification	Ordination	
Real	+	+	+	+	+	
Data	+	+	+	+	+	
Derived variables		+				
Resemblance		+	+	+	+	
Classification				+		
Ordination					+	

4.1 *Dependent processes in data space*

The lowest level at which the effect of sampling design may be seen is represented by the raw scores for the variables in consideration. In the study of plant communities the most commonly used variables are cover, number of individuals (count), and biomass (yield) within sample plots. Cover is defined either as the area of the perpendicular projection of all parts of individuals of the same species onto the ground (aerial cover) or the actual basal area occupied by the plants. Absolute cover is seldom used, it is rather relativized according to the area of the sampling unit (percentage cover). The counts may also be divided by plot size to give density, the number of individuals per unit area (CURTIS and MCINTOSH 1950). The subject of this section is to study changes of these variables in course of a primary process in real space. Other vegetation variables, for example those frequently used in pattern studies (inter-plant distances, etc.), are not considered here. It is noted that the total cover and biomass of several species are derived variables and will be discussed in 4.2.1.

Most of the published work about the influence of sampling on data is devoted to the question of plot size. Early in this century, RAUNKIAER (1918) suggested that for the determination of cover the size of plots is theoretically immaterial. He considered plot size as a practical issue saying in as much as "*it is easier to estimate correctly on a small than a large area*". PENFOUND (1945), using artificial communities of randomly arranged cards, showed that percentage cover and density did not vary with plot size. In a sand-steppe community, MOLNÁR and NÓSEK (1980) examined the changes of living phytomass, number of individuals and absolute cover of five species over plot size. Strongly significant linear correlations were found between each variable and the area in all cases. It might be then concluded that a homogeneous vegetation stand may be sampled by any plot size, but obviously this is not the case. Since the data are usually subjected to further analysis, the sample has to satisfy certain optimality criteria. Depending on the goal of the survey, the between and within plot variances should be minimized or maximized, costs have to be reduced or other conditions met. Therefore, the level of raw data is almost always insufficient to make correct decisions even if the assumption of linearity between variables and plot size generally existed. This problem brings up the importance of derived variables.

4.2 *Dependent processes in the space of derived variables*

A derived variable is usually defined as an expression combining two or more independent and direct measurements on biological material (SOKAL and ROHLF 1973). GREEN (1979) makes a distinction between four types: ratios, percentages, indices, rates, and he provides examples applicable to environmental studies. However, for the present purposes it seems reasonable to give a broader definition. Any variable which reflects meaningful properties of the data other than those expressed by the measurements themselves will be considered as a derived variable. Most of these are derived from the raw data based on arithmetic operations. The resemblance coefficients are excluded at this time, since they will be discussed separately (4.3), although in special cases derived variables may also be defined based on resemblance matrices (i.e., expected resemblance, section 4.2.10). As a consequence of this definition, total cover, number of species, mean and variance of population parameters usually not considered as derived variables will fall into this category.

4.2.1 *Total cover, number of individuals, and phytomass*

If interest lies in the determination of total cover, total count, and biomass of all species or a particular subset of species, and data have been recorded at the species level, the necessary information is obtained by summing the corresponding raw scores pertaining to

the same plot. There is at least one report (MOLNÁR and NOSEK 1979) suggesting that in a homogeneous vegetation stand the relationship between these derived variables and plot size is significantly linear. In other words, if these variables are relativized to unit area, no significant change will be seen. Similar conclusions were drawn by VAN DYNE et al. (1963) for production estimates of various plant groups.

4.2.2 *The precision of estimates*

Determination of cover, biomass per unit area and density for the whole population of a particular species is important for plant ecology and related fields (e.g., production biology, forestry, agronomy, resource inventories, etc.). These variables, especially biomass, are frequently measured for an entire species group or for all the species. The precision of estimates derived from the sample is greatly influenced by the sampling design. The optimization of sampling to achieve high precision is commonly carried out by a statistical analysis of spatial processes. Precision is expressed in terms of variance, sampling variance of the mean or its square root (standard error of the mean) and the coefficient of variation. An "ideal" design would give data with the least variance and would result in an unbiased estimate of the true population mean at the lowest costs. Of course, all requirements cannot be met at the same time in practice.

Intuitively, the precision of estimates at a fixed sample size is inversely related to plot size since higher proportion of the whole population is sampled by larger plots than by smaller ones. However, this relationship is also affected by the spatial pattern of species under investigation. If the population analyzed shows clumped distribution, the variance between plots would increase, and not decrease, over plot size up to the point where the plots approximately equal the mean area of clumps. Further enlargement of plots is not a guarantee of increasing precision because the variance also depends on the pattern of the clumps themselves (cf. PIELOU 1977). Therefore, one might be tempted to accept the more precise estimates obtained by very small plots, as suggested by GREEN (1979): "*the best sampling unit size generally is the smallest possible one, when sampling a given total area for an organism that has an aggregated spatial distribution*". However, if the plots are small the sample size has to be increased to reach an acceptable sampling intensity. This problem illustrates that two potential objectives of vegetation surveys, namely the precision of estimates and the detection of pattern are conflicting. Due to the aggregated spatial distribution, abnormally high variances in yield were reported by VAN DYNE et al. (1963) from sampling experiments with grasses. On the other hand, the increase of precision as a result of plot enlargement is observed in many studies for other species. VAN DYNE et al. (1963), WIEGERT (1967), THOMAS and ABOU-EL-FITTOUH (1968) and RAM BABU et al. (1981), among others, provide evidence for biomass data in economically important natural communities. FREESE (1961), O'REGAN and PALLEY (1965), O'REGAN et al. (1973), TAAFFE (1979), FOWLER and DAVIS (1979), and KULOW (1966) have drawn similar conclusions about the precision of various forest estimates (number of trees, basal area, volume of timber, etc.). According to FREESE (1961), the product of the square root of plot size and the coefficient of variation of timber measurements is approximately constant. This experimental formula gives a good fit to actual data as confirmed by TAAFFE (1979) and ZEIDE (1980). For density data O'REGAN et al. (1973) found another formula suggesting that the logarithm of variance is a quadratic function of the logarithm of plot size.

It was already mentioned in 3.1.2 that plot shape may significantly influence the precision of cover, density and biomass estimates. The problem of appropriate shape has long been a subject for investigations by agronomists and it was early recognized that long narrow units are effective in reducing the between-plot variance of yield measurements

(KALAMAKAR 1932, JUSTESEN 1932). CLAPHAM (1932) obtained the same type of results for counts of five herbaceous species in natural vegetation. HASEL (1938), BORMANN (1953) and FREESE (1961) reported that long plots were superior to squares or circles in reducing variance in forest measurements. PECHANEC and STEWART (1940) and VAN DYNE et al. (1963) evaluated the influence of shape upon plant production in different steppe communities with similar conclusions. The advantage of elongated units over isodiametric ones is due to the fact that the chance of crossing any vegetational or environmental boundaries is higher for the long sample plots. That is, between plot variances are minimized by maximizing within plot heterogeneity. Consequently, in analyzing homogeneous vegetation plot shape seems less critical, although CLAPHAM (1932) recommended the exclusive use of rectangles in Phytosociology stressing that "the soil characteristics which affect plant growth cannot be absolutely constant over an area, however small it may be". Reports on the inconsistency of plot shape with respect to precision, however, do not support the view that this factor is universally important in estimation (e.g., KULOW 1966, RAM BABU et al. 1981). Moreover, if the objective is not estimation, the use of long plots is not necessarily preferable (e.g., in the analysis of interspecific correlations, 4.3.1).

The effect of sample size upon the precision of estimates is mathematically more tractable than any other characteristic of sampling. Therefore, the intuitive feeling that the larger the sample size the higher is the precision may be easily tested. Let N denote the number of non-overlapping sample plots of a given size and shape completely covering the study area. Supposing that the sampling is free from measurement errors, the mean of a sample consisting of these N plots would be completely accurate. Let n be the number of non-overlapping and random plots actually used, and S_X^2 be the variance between plots. Then the unbiased estimate of the standard error of the mean is obtained by the formula,

$$S_{\bar{X}} = \sqrt{\frac{S_X^2}{n} \cdot \frac{\sqrt{N-n}}{N-1}} \quad (8)$$

(refer to any statistical text, e.g., FREUND 1967). $S_{\bar{X}}$ should be minimized to achieve high precision. It is seen that if N is very large, the finite population correction factor, $\frac{\sqrt{N-n}}{N-1}$, becomes negligible and the value of n is decisive in determining precision. It is also obvious that decreases in standard error are proportional to the square root of n , therefore it is not worth increasing the sample size beyond a reasonable limit. Given a desired accuracy of the mean, a specified confidence interval and an estimate of S_X^2 based on a pilot survey, the necessary sample size can be determined (see COCHRAN 1963, SAMPFORD 1962, GREEN 1979, for details and examples). The effect of sample size on the precision of yield estimates was studied by THOMAS and ABOU-EL-FITOUH (1968) who suggested that increasing replication number reduces error more rapidly than increasing plot size. ROGERS (1980) made a similar survey on percent cover data and showed that the coefficient of variation was rapidly decreasing versus sample size if the plots were randomly arranged.

The relationship between the precision of estimates and the arrangement of sample plots depends on the pattern of variability within the study area. No doubt that in homogeneous vegetation simple random sampling is the obvious strategy. If heterogeneity is indicated by pilot sampling, simple random sampling is considered unsatisfactory, since it is unlikely that the different zones or strata will be proportionally represented in a random sample. Systematic sampling was first thought to be applicable to overcome this difficulty and its superiority was reported from forest studies by HASEL (1938) and FINNEY (1948). However, other reports comparing systematic and simple random sampling showed that there is little or no gain in precision (BOURDEAU 1953, KULOW 1966). In populations with periodic variation the precision

of systematic sampling is sensitive to the spacing of plots. If the sampling interval coincides with the scale of pattern the resulting variance will be extremely low because the other regions are excluded from the sample (cf. COCHRAN 1963). A smaller or larger spacing, on the other hand, may lead to much higher variance than simple random sampling, as reported by FINNEY (1950). Obviously, systematic sampling cannot be used to derive unbiased estimates of the sampling variance, and comparison to other designs is not valid on this basis. COCHRAN (1963) describes a method for objective evaluation and shows that if the variance between units is greater than the population variance, systematic sampling is expected to be more precise than is simple random sampling. This is the case if the population exhibits a linear trend, which should be very rare in natural vegetation. Therefore, another alternative to simple random sampling, namely stratified random sampling, may be more useful in diminishing sampling error. Its increased precision against the simple randomized arrangement has been demonstrated by PECHANEC and STEWART (1940) for yield and BOURDEAU (1953) for density and basal area of trees. The reason for higher precision is that systematic sampling ensures a relatively even distribution of plots within the study area. The gain in precision obtained by proportional allocation of plots depends on the difference between the stratal means: the higher the differences, the more precise is the stratified design compared to the simple random design. By definition, the optimal allocation strategy yields the highest precision, since the number of plots allocated within each stratum is determined in order to decrease the sampling error within strata. A theoretical advantage of stratified sampling is that estimates of sampling variance are valid, but difficulties may arise in delimiting vegetation strata (cf. KENKEL 1981).

The effect of spacing upon the precision of estimates is analogous to the problem of plot size/variance relationships in that the objectives of statistical estimation and pattern detection are conflicting. Whereas plot size seems equally important from both points of view, spacing received more attention in a pattern analytical context (see 4.2.11). The influence of spacing on the variance of systematically arranged plots was already discussed. The author is not aware of work examining other aspects of spacing/precision relationships.

The uniform orientation of long plots appears an important factor affecting the variance if they are arranged along a striking environmental gradient and the vegetation is assumed to be homogeneous perpendicular to it. LAFRANCE (1972) examined this problem using computer-simulated data, but the results did not verify the above statement. LAFRANCE attributed this to the relatively low length/width ratio of his plots and, following BORMANN (1953), recommended the use of more elongated units to improve precision.

4.2.3 *The accuracy of estimates*

In fact, the goal of any attempt to increase precision is to increase the accuracy of percent cover, density and biomass estimates. The reason that different sampling designs are evaluated on the basis of precision lies in the fact that the population means are not known. The "true" mean of a variable can be determined only by complete enumeration which is very expensive if not impossible in a large study area. Further difficulty is that several samples from the same area cannot be taken independently in the field without the effect of drastic disturbance in the vegetation. These problems render the examination of the accuracy of yield estimates all but impossible. Vegetation mapping and computer simulation of spatial pattern, however, may be used to investigate the accuracy of cover and density estimates obtained by different designs. It is always assumed in surveys that the determination of plant coordinates and cover is free from measurement error and the mean obtained for the whole area is "true". KULOW (1966) used maps of three different forest stands with information on tree diameter. The basal area of trees per unit area was the variable used for

comparison. Plot shape was of no consequence, and the accuracy of random and systematic sampling varied with forest type. The enlargement of plots was very efficient in reducing the deviation of estimates from the population mean. In a very different approach, LAFRANCE (1972) employed different plot shapes to estimate the mean location of species along a gradient in an artificial population. Long rectangles, both in parallel and perpendicular orientation, gave the most accurate estimates for species with aggregated pattern. For randomly arranged species, as might be expected, the plot shape was immaterial and no consistent relationship to accuracy was indicated.

4.2.4 *The distribution of variables*

Many statistical tests and several methods of multivariate analysis are based on the assumption that the distribution of variables is normal. Symmetric, bell-shaped distribution curves are usually accepted as good approximations to normality. For randomly dispersed populations, the symmetry of the curve of the number of quadrats versus the number of individuals per quadrat is affected by the mean which is in turn a function of plot size (cf. COTTAM et al. 1953). In aggregated populations this relationship becomes more complex. As a rule of thumb to be followed in deciding plot size, GREIG-SMITH (1964) suggested to select the smallest area which will not give more blank plots than plots with a single individual.

The distribution of cover data was first examined by WEST (1937) using basal area percentages for several species and with different quadrat sizes. The raw scores were grouped into equal cover classes (as described in 3.2.3), and the number of sampling units was plotted against percentage area for various species and plot sizes. The curves indicated that the distribution of cover class data of a common species was skewed at small plot size, whereas at large plot sizes the distribution appeared to be normal. This experiment was repeated by RICE (1967) with North American species. The skewness of the curves also disappeared and the distribution became symmetric as the plots were enlarged. However, this trend was observed for only three common species. Despite its importance, the normality of variables is hardly ever investigated in recent surveys.

4.2.5 *Frequency*

The frequency of a species is traditionally defined in Phytosociology as the number of plots in which it occurs, expressed as a percentage of the total number of plots examined (CURTIS and McINTOSH 1950). In other words, it is an estimate of occupancy expected if a quadrat of the given size is placed randomly within the study site. To avoid confusions, it is noted that in statistics frequency is understood as the observed number of outcomes of a given event (in this case the number of plots in which a species occurs), but now the phytosociological terminology will be used. This concept is the source of much controversy between traditional phytosociological schools. DU RIETZ (1921) claimed that the number of species pertaining to arbitrarily established frequency classes (0–20%, 21–40%, 41–60%, 61–80%, 81–100%; the frequency distribution of RAUNKIAER) is a characteristic of a homogeneous vegetation, and plot sizes larger than the “*minimal area*” have virtually no influence on this distribution, at least within a considerably broad range (cf. BLACKMAN 1935). It was an apparent challenge for expression of opposing views since it is the homogeneity which ensures that the increase of sampling area will lead to increasing frequencies. Then, not surprisingly, WEAVER and CLEMENTS (1938) commented that “*frequency depends so much upon the size of the quadrat employed as to be without value by itself*”. PENFOUND (1945) concluded from the study of artificial populations that frequencies and the number of species in the frequency classes of RAUNKIAER vary with quadrat size. CURTIS and McINTOSH (1950) examined the relationship among frequency, density and plot size in a simulated community and demon-

strated that frequency increases more rapidly with increasing quadrat size when densities are great than when densities are low. The shape of RAUNKIAER's frequency distribution curve was markedly affected by plot size. The authors concluded that a RAUNKIAER distribution is of no value in indicating homogeneity of a community, and may be interpreted only with reference to quadrat size and density. GREIG-SMITH (1964) pointed out that frequency also depends on the spatial pattern of species. He examined the theoretical influence of sample size on frequency estimates, and demonstrated that the confidence limits for frequency values are too broad if the estimate is drawn from a small sample. As an obvious remedy to this, the use of very large sample size was recommended. In view of the contemporary vegetation studies the debate over the validity of frequencies seems fairly sterile, although species frequency estimates are still used as input data for information analysis (e.g., WILLIAMS et al. 1969, JUHÁSZ-NAGY 1976, etc.), and not without success.

A relative frequency of a species is obtained by dividing its frequency by the total of the frequency values for all species. CURTIS and MCINTOSH (1950) illustrated that in a homogeneous simulated population the relative frequencies approach the value of $100/n$ (n is the number of species) as plot size increases. In fact, this result was obtained independently in the florula diversity surveys of field data by JUHÁSZ-NAGY and PODANI (1983), where zero florula diversity implies equal relative frequencies. This is the result if all species are present in all sample plots (see 4.2.9).

4.2.6 Relative density

This quantity is the percentage of the total number of individuals of a species relative to the total of all species. By definition, this is the estimated probability that a randomly selected individual belongs to this species. CURTIS and MCINTOSH (1950) showed that relative density is largely independent of plot size. The accuracy of the estimate in random populations, however, proved to be subject to plot size and sample size in another study of simulated populations (COTTAM et al. 1953). It was demonstrated that accuracy is in fact a function of the number of individuals enumerated in the sample. This is of primary consequence in the species/individual diversity surveys to be discussed in 4.2.8.

4.2.7 The number of species

The relationship between species number and area has been the subject of numerous studies in various fields of biology. In Phytosociology, fundamental ideas of different schools are based on the examination of species/area curves. Before turning to these concepts it is mentioned that a potential source of difference in interpretations is that there is no unified data collection method for species/area curves. Two basic strategies may be used (cf. GREIG-SMITH 1964, PIELOU 1977):

1. A "true" species/area curve is derived from a series of samples in which every sample plot is assumed to be independently allocated, regardless the size of the others. Data for each plot size are then screened and the mean number of species per quadrat is calculated. The estimates of the mean are unbiased since the design is completely random. This ideal strategy seldom occurs in actual studies (e.g., HALL and OKALI 1979).

2. Cumulative curves ("collector's curves", PIELOU 1977) are obtained based on either a nested design or by increasing sample size. In this case data for different plot sizes are not independent, a rare species occurring in one of the smallest or the first plots by chance will necessarily be present in all samples. Although interdependence may be a drawback from a statistical point of view, this method is generally used. Most functions of the type,

$$S = a + b \log x \quad (9)$$

where S is the number of species and x denotes area, suggested by several authors (see GREIG-SMITH 1964, for details and references) are in fact approximations to cumulative curves. Other formulae (e.g., KILBURN's function, see below) were also inferred from a nested design. PIELOU (1977) discussed the relationship between species-abundance distributions and cumulative species/area curves with relevant conclusions regarding the homogeneity of the community sampled. However, as PIELOU herself discovered, species/area curves from plant communities cannot be interpreted in practice. The assumptions of the existence of a log-series or a negative binomial distribution for plants are usually not satisfied (PIELOU 1977). Further difficulty is that plant individuals cannot be distinguished in case of many species (e.g., grasses).

The relation of species number to sampled area has been investigated with much effort but apparently with little success. There were many attempts to develop a method capable of selecting a particular point on the curve or on the area axis and thereby determine a characteristic area. The minimal area concept by BRAUN-BLANQUET (1951) relies on the statement that the $S : x$ curves level off at a given plot size, but this is not always expected to happen in view of more recent studies (cf. KILBURN 1966). HOPKINS (1955) examined several communities and found that the $S : x$ curve was roughly linear for large quadrats. He suggested to use the interception of the projected linear segment and the area axis to determine an "objectively defined area of the community". KILBURN (1966) proposed a "power function" which appeared to give a good fit to the actual data and suggested its first derivative as an objective technique to find optimum plot size. Other, more or less arbitrary techniques are reviewed by CAIN and CASTRO (1959), GREIG-SMITH (1964) and bibliographed by TÜXEN (1970). There is no doubt that species/area curves may reflect important intrinsic properties of plant communities but it appears that their use to find characteristic areas for further analysis is not theoretically justifiable. The "objective" area obtained by HOPKINS' method was always too small to be biologically significant whereas KILBURN's technique involves an arbitrary selection of "appropriate slope values". However, the fundamental problem with the species/area curve as a basis for optimum plot size determination is that the optimum differs with the purpose of the study. It is sufficient to recall the contradictory objectives of estimation and typification. The optimum has to be found by a technique compatible with subsequent aims and operations. Manipulations with species/area curves cannot satisfy all requirements.

The shape and arrangement of sample plots have been little investigated in this context. These factors have probably no practical influence on the observed species number even in a species rich community if a reasonably large plot and sample size is used. It may be assumed, that very long but small plots are more likely to contain more species than isodiametric ones of the same size (cf. NOSEK 1976). If the sample size is small, the systematic and stratified random designs are expected to indicate more species than a simple random arrangement.

4.2.8 Species/individual diversity and evenness

In the past two decades diversity indices were among the most widely used mathematical tools applied to ecological studies. Whereas the discussion of the many available formulae and the criticism of the popular concepts of diversity is beyond the scope of the present work, reference is made to PEET (1974), PIELOU (1975) and GREEN (1979) for exhaustive treatments. The subsequent text discusses the spatial aspects in the way that these relate to the measurement of diversity and evenness.

The diversity of a multi-species population is a function of species number and the number of individuals belonging to each species. In a fully censused population the diversity is exactly known, supposing that there is no error in the counts. If complete enumeration is

impossible, as it is in most surveys, the diversity is estimated from a sample. Clearly, the estimation of the number of species and the relative density of each is improved by increasing the number of individuals in the sample (cf. 4.2.6-7), therefore the same effect is expected for diversity. Consequently, the increase of plot and/or sample size is an efficient way to improve precision. This is illustrated by WOLDA's (1981) simulated experiment in which the standard deviation of diversity estimates obtained by replicate samples decreased as sample size increased. The shape and arrangement of plots do not appear important unless the total sampled area is small.

Empirical results (e.g., HEYER and BERVEN 1973, NOSEK 1976, FEKETE and KOVÁCS 1978) suggest that the curve of diversity estimates versus sampled area (sample or plot size) first increases rapidly, but not necessarily monotonically, as it is approaching the "true" diversity of the community. At large sampled areas the estimates are slightly oscillating and no upward trend can be seen. Diversity curve may serve as a basis for estimating diversity in terms of SHANNON's entropy, as described by PIELOU (1975). The relationship among sample size, diversity, and underlying spatial arrangements of species was examined by KOBAYASHI (1981), using simulated data. He showed that some of the parametric diversity indices were markedly influenced by sample size depending on the spatial distribution. Shannon's entropy seemed to be the least dependent on sample size regardless the type of distribution.

The concept of species evenness is closely related to diversity. By definition, the evenness of a community is maximal if all species are equal in abundance and minimal if all but one species are represented by a single individual. Simulations and examinations of real data (e.g., NOSEK 1976, KOBAYASHI 1981) demonstrated that species evenness tends to decrease as the sampled area increases. This is purely a mathematical fact without a biological basis, if the number of species increases at the same time (cf. PIELOU 1975). In other words, evenness estimates derived from a sample are usually overestimates of the "true" population value. Therefore species evenness appears of no theoretical and practical significance unless the species number of the community is known (PEET 1974, PIELOU 1975). Contrary to animal populations, the determination of species number is not necessarily an impossible task in terrestrial plant communities. The real difficulty with diversity and evenness measurements in vegetation studies is that plant individuals are not always distinguishable.

4.2.9 *Supraindividual diversity*

In the analysis of the vegetation, a fundamental problem is to determine the spatial variation of a community over an area or an environmental gradient. Species/individual diversity is completely uninformative in this respect for obvious reasons. A more adequate approach is due to JUHÁSZ-NAGY (1967, 1976) who suggested the use of SHANNON's entropy function to measure the diversity of a sample on the basis of the species composition of each plot. Since the basic unit is not the individual but a collection of individuals (i.e., those present in a quadrat) this kind of diversity is generally termed supraindividual diversity (in case of plant communities it is called florula diversity, cf. JUHÁSZ-NAGY and PODANI 1983). A substantial difference between this and species/individual diversity is that the spatial dependence of supraindividual diversity is not simply the matter of estimation.

The florula diversity of a sample consisting of m plots is given by,

$$m\hat{H}(A, B, \dots, S) = m \log m - \sum_{k=1}^{2^s} f_k \log f_k \quad (10)$$

where A, B, \dots, S symbolize species, s is the number of species and f_k is the number of quadrats with the k th species combination (florula). Note that 2^s different species combinations are possible including the one containing no species ("empty" florula). It is assumed that the plots follow a random arrangement.

There is evidence in the literature (see the references made in this section) that quantity (10) first increases and then decreases if plot size is increased. The plot size where the maximum is obtained is characteristic of the community and is termed the "maximum area" (JUHÁSZ-NAGY and PODANI 1983). The maximum does not necessarily imply that all the 2^S potential florulas may be manifested in the sample, since most theoretical combinations cannot occur due to the non-random ground pattern of species. By the same reasoning, some particular combinations will be more likely to occur than others. At a much larger plot size florula diversity becomes 0, this is the "minimum area" in terms of florula diversity.

It can be shown that (10) is in fact the joint entropy of the species:

$$m\hat{H}(A, B, \dots, S) = m\hat{H}(L) - m\hat{I}(A, B, \dots, S) \quad (11)$$

where $m\hat{H}(L)$ denotes the estimate of the local distinctiveness of species ("total information", ORLÓCI 1969, SNEATH and SOKAL 1973, "information content", WILLIAMS et al. 1969) and $m\hat{I}(A, B, \dots, S)$ is the estimate of association (or mutual information) among species. These functions also have maxima, indicating two other characteristic areas, namely the compensatory area and the area of maximum associatum, respectively (see JUHÁSZ-NAGY and PODANI 1983, for more details).

Supraindividual diversity and related characteristic functions have also been applied to crustacean plankton assemblages ("faunal diversity", DÉVAI et al. 1971, JUHÁSZ-NAGY et al. 1973). The changes of these functions during succession were studied by JUHÁSZ-NAGY and PODANI (1983). Local distinctiveness only was measured by WILLIAMS et al. (1969) at various clump sizes in multiple nearest neighbor analysis (3.1.5). The extension of these studies to many other community types would be desirable in order to provide further information on characteristic functions as to their potential use in optimizing plot size.

4.2.10 Expected resemblance

Additional information about spatial variability in a community is conveyed by the expected resemblance (similarity or dissimilarity between two random plots). An expectation can be obtained in terms of the non-diagonal elements (d_{jk}) in a resemblance matrix,

$$E(d_{jk}) = 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m d_{jk} / (m^2 - m) \quad (12)$$

Since the values of the resemblance matrix are not independent, one satisfactory but laborious procedure to estimate the expectation would be repeated sampling based on random pairs of quadrats. Clearly, the expected value is subject to change as plot size is modified. Several past attempts have been made to find characteristic areas on the basis of the spatial dependence of expected resemblance. GOUNOT and CALLÉJA (1962) found that the mean similarity of sample plots first rapidly and the slowly increases as plot size increases, whereas the variance of similarities is decreasing. They defined the minimal area of the community as the smallest plot size at which the expectation significantly exceeds a "given value". GOUNOT (1969) claimed that once the plot size corresponding to this minimal area is reached, the mean similarity cannot rise any more. MORAVEC (1973) applied SØRENSEN'S coefficient to find the minimal area of different communities with varying results. For meadow data, the curve of mean similarity versus plot size appeared asymptotic, at least within the range of the plot sizes used. In two cases, however, the expectations did not show consistent trend, the slight changes were probably due to random variation. For one community, the results were even more different, and the curve showed two striking peaks. The highest value was obtained at

a very small plot size (0.25 m²), the other maximum occurred at 100 m². These two maxima clearly indicated two different scales of pattern, corresponding to the understory vegetation and the tree-layer, respectively. In fact, MORAVEC's study did not verify GOUNOT's general statement and it appears that the relationship between plot size and expected similarity is a more complex one than earlier thought. This statement is supported by DIETVORST et al. (1982) who suggested to consider not only binary data but also quantitative scores and dominance relations in determining minimal area. However, the concept of an arbitrary lower bound of similarity was adopted in two of the three different situations they recognized.

At first glance the approach of expected resemblance based on the similarity of binary data, and the concept of florula diversity seem contradictory. The definition of minimum area (4.2.9) implies that the expected similarity is maximum at the minimum area, but this plot size was not reached in the studies by GOUNOT, CALLÉJA and MORAVEC. It is obvious, then, that the "minimal area" in terms of expected resemblance is much smaller than the minimum area of florula diversity. The area of maximum florula diversity, however, does not necessarily imply that the expected similarity is minimum at this point, since florula diversity does not reflect direct inter-plot similarities. The simultaneous application of the two methods to the same set of data would be desirable to clarify the relationship between them.

4.2.11 Pattern

The previous two sections have already touched on the problem of spatial pattern. The objective in examining florula diversity and expected resemblance is in fact the analysis of multi-species pattern on the basis of the qualitative description of vegetational plots. This section is devoted to a much more elaborated topic, the study of the spatial pattern of a single species. There are two fundamentally different approaches to this: the quadrat method and the distance method of plotless sampling. The subsequent discussions are primarily concerned with the first approach.

The analysis of plant pattern by the quadrat method has been the subject of many investigations in past decades. The fundamental question addressed is whether the ground dispersion of individuals can be regarded as random, or if not, aggregated (contagious, cf. PIELOU 1977, p. 117) or regular. Various measures of the departure from randomness have been suggested, these are discussed in detail by GREIG-SMITH (1964) and PIELOU (1977). ROGERS (1974) provides a monograph on the statistical analysis of spatial dispersion by quadrat methods.

In a randomly dispersed population the expected underlying distribution of the number of individuals per plot is Poisson, regardless the plot size. Consequently, the suitably chosen measures (e.g., goodness of fit to the Poisson series, variance : mean ratio) are also independent of plot size. If the measures indicate no significant departure from the random dispersion over a wide range of quadrat sizes, the hypothesis of randomness may be accepted. In natural vegetation, however, randomness seldom occurs (cf. GOODALL 1974), although it is commonly assumed for rare species. As GREIG-SMITH (1964) points out, this conclusion may stem in part from the inappropriateness of quadrat and sample sizes used and the results should not be accepted without reservations.

If the spatial dispersion of plant individuals is non-random, the measures of the departure from randomness will not be independent of plot size (although PIELOU 1977, lists two special types of pattern for which this statement does not generally hold true). In a regularly dispersed population very small plots may fail to detect non-randomness, but this problem may be easily overcome. The variance : mean ratio tends to detect more apparent regularity as plot size increases (cf. CURTIS and MCINTOSH 1950, GREIG-SMITH 1964) if the spatial dispersion is actually regular. It is to be noted, however, that this question is rather

theoretical, since there is no good evidence about the existence of two-dimensional regular pattern in natural populations. The dispersion of desert shrubs was generally thought to be approximately regular, but a recent survey (EBERT and McMASTER 1981) showed that this conclusion is false.

The overwhelming majority of plant species show contagious spatial distribution, where the individuals are aggregated into clumps as a result of vegetative propagation, environmental heterogeneities or spatial exclusion by other species. In this case the sampling unit size is most crucial, the dispersion may appear random at very small and regular at large quadrats, whereas significant aggregation is detected at intermediate sizes (see the example by KERSHAW 1973, p. 136). Consequently, it seems that the most admissible strategy in the analysis of aggregated populations is a sampling with various plot sizes followed by the examination of dependent processes. The various indices, as mere descriptors of pattern, may be plotted against quadrat size. The shape of the resulting graph provides some useful information about the spatial dispersion of individuals. The point(s) corresponding to the scale of pattern (mean area of clumps, etc.) are indicated by peak(s) which may be readily recognized. The troughs on the graph may be as informative as the peaks, they are indicators of regularity (see e.g., RIPLEY 1978). Theoretically, the use of independent random plots would represent the most satisfactory sampling design, since the results pertaining to each plot size could be compared by statistical tests. The second best strategy would be random sampling by nested plots (Fig. 2a-d), although the possibility of testing significance is lost. In practice, however, the use of grids and transects of contiguous quadrats has become widespread in ecology, since randomization and, especially, repeated sampling with various plot sizes is exceedingly time-consuming and the vegetation becomes more and more trampled as the sampling proceeds. The advantage of systematic sampling lies in the fact that counts of the number of individuals, cover estimates or the weight of clippings have to be determined only one time for each unit, and increasing "quadrat" sizes are obtained by combining these data in some regular way.

A technique of pattern analysis utilizing the information collected from a grid of contiguous quadrats was suggested by GREIG-SMITH (1952) and is commonly referred to as "grid analysis". The fusion strategy of this method to obtain enlarged "quadrats" was already discussed in 3.1.6. The mean square among the m grid cells is examined by a nested analysis of variance. The total sum of squares is apportioned into independent components that derive from the difference between each pair of quadrats within the two-unit blocks, from the difference between each pair of oblong, two-celled blocks within the four-unit blocks, and so forth. The sum of squares specific to blocks consisting of r units is

$$SSQ_r = \frac{1}{r} \sum_{i=1}^{m/r} x_{i(r)}^2 - \frac{1}{2r} \sum_{i=1}^{m/2r} x_{i(2r)}^2 \quad (13)$$

where $x_{i(r)}$ and $x_{i(2r)}$ are the number of individuals in block i of the respective size. The corresponding mean square is

$$MS_r = 2rSSQ_r/m. \quad (14)$$

Assuming a random distribution of individuals, each mean square is an estimate of the variance of the Poisson distribution for the counts in the grid units. Then, the ratio of mean square/overall mean is approximately 1 for each block size. If the dispersion of individuals is aggregated, MS and, consequently, the variance/overall mean ratio will increase until a block size is reached which approximately equals the mean area of the clumps. If the clumps themselves are regularly dispersed, MS will fall as the blocks are further enlarged, but if the clumps show a random arrangement the mean square will remain about the same level. Two or more peaks

may occur if the individuals are dispersed in a hierarchy of clumps. The same technique may be applied to cover data (KERSHAW 1957) or biomass (MORTON 1974).

The increase of plot size in grid analysis implies a complex primary process (cf. 3.1.6), and the simultaneous change of shape and size may give rise to confounding effects in dependent processes. It is sometimes observed that the graph of MS versus block size has a sawtoothed shape because oblong blocks consistently give mean squares less than those of the neighboring square blocks in the sequence of sizes (PIELOU 1977). Due to the alternating shape of blocks, the center-to-center distance between them is not changed when square blocks are combined into oblong ones, and this fact alone may lead to increasing mean squares in a non-uniform population (GOODALL 1963). Therefore, instead of fusing adjacent blocks, GOODALL (1974) suggested to use the center-to-center distance as conditioning parameter in pattern analysis. Random pairs of quadrats with various distances apart are picked up from the grid without replacement and the specific variances are plotted against quadrat spacing. This technique overcomes the difficulty of the dependence of successive mean squares, and standard statistical methods can be used to test whether a significant difference exists between two spacings.

Other modifications to GREIG-SMITH's grid analysis are due to ORLÓCI (1971) and ZAHL (1974). ORLÓCI used the minimum discrimination information statistic to measure the departure from a perfect regularity (see also FEOLI et al. 1982). Confidence limits were experimentally derived to examine whether a pattern is contagious or random. Zahl adopted SCHAFFÉ's S-method of multiple comparisons in estimating the size of patches. This technique requires the use of overlapping blocks. Spatial autocorrelation (CLIFF and ORD 1981) was first applied to the analysis of plant pattern by CHESSEL (1981) and GLOAGUEN and GAUTIER (1981).

KERSHAW (1957) recognized that the orientation of the oblong blocks from the grid units may considerably influence the shape of the graph of mean squares, if the mosaic patches are not isodiametric. Despite this problem, long transects of quadrats are more commonly used in pattern analysis than two-dimensional grids. Clearly, the results obtained by these two different means are not expected to be the same, since the transect method is capable of detecting only one-dimensional spatial pattern. The use of transects is therefore recommended only in case if there is a good reason in selecting a particular orientation. GREIG-SMITH's or ORLÓCI's technique can be used, but they are probably very sensitive to the starting position of the transect (cf. USHER 1969). To avoid this, many variants of the original method have been developed (HILL 1973a, LUDWIG and GOODALL 1978, FEOLI et al. 1982). The paired quadrat-variance method suggested by LUDWIG and GOODALL is based on all possible pairs of quadrats to estimate the variance at each spacing. This method is actually equivalent to the correlogram used in time-series analysis. A recent comparison and review of the quadrat-variance techniques is due to LUDWIG (1979). Repetitive one-dimensional patterns can be detected by the paired-quadrat variance method or spectral analysis (e.g., HILL 1973a, USHER 1975, RIPLEY 1978).

4.3 *Dependent processes in resemblance space*

Changes in resemblance space are examined considering a primary process defined in either real or in data space. The study may be restricted to a pair of objects (species or quadrats) in which case the dependence of resemblance upon the primary process is simply demonstrated by a profile diagram (e.g., AUSTIN 1968). If all possible pairs of objects are considered, the illustration becomes more difficult and the use of plexus diagrams may be appropriate (MCINTOSH 1973). A plexus diagram is undirected and coloured, but not necessarily a connected graph in which vertices represent objects (usually species, e.g., AGNEW 1961). Two vertices are linked by an edge of colour c_i if the resemblance between the corresponding

objects exceeds the threshold t or, if an appropriate statistic is significant at probability level t . A series of such plexus diagrams may sufficiently illustrate an association process of species in resemblance space (e.g., FEKETE and SZUJKÓ-LACZA 1973, MATTHEWS 1978). A more adequate approach is the comparison of resemblance matrices by elements to express the similarity of two matrices by a single number. The use of product-moment correlation coefficient, as a similarity index and not a statistic, was suggested by SOKAL and ROHLF (1962) for this purpose. Since then matrix correlation has been widely used by taxonomists (cf. ROHLF and SOKAL 1981) but the number of ecological applications is limited (e.g. ORLÓCI and MUKKATTU 1973, see below).

4.3.1 *Resemblance between species*

It is often emphasized by plant ecologists that interspecific association is greatly affected by the size of sample plots. GREIG-SMITH (1964) warned that the interpretation of interspecific associations is difficult if the analysis is based on a single plot size. It seems, however, that the use of various plot sizes does not necessarily facilitate the assessment of interspecific associations. GREIG-SMITH himself pointed out that there are at least four possible causes of the dependence of associations upon plot size. Accordingly, it is not always known what causes can be accounted for by the changes.

As yet, plot shape and arrangement have not been investigated in this respect. It is easily seen, however, that isodiametric units give the most acceptable results since elongated units may indicate positive correlation between two species that are actually far apart in the field. In a heterogeneous area stratified random sampling is expected to give more reliable indication of interspecific associations than a simple random design. In this case the proportional allocation of plots is an extremely important condition. Intensive sampling within a given stratum, from which the species considered are missing, may indicate abnormally high positive associations (BRAY 1956, gave an example).

An interesting approach by NOY-MEIR (1971) should be mentioned. He explored the effect of quadrat size on the ratio of qualitative and quantitative components in the correlation matrix of species. It was found that the larger the quadrat size, the more important is the quantitative component in the resemblance matrix.

4.3.2 *Resemblance of quadrats*

ORLÓCI and MUKKATTU (1973) presented two examples for the use of matrix correlation in ecology. They investigated the effect of species number and individual count limit (3.2.3) on the resemblance structure of quadrats. Resemblance matrices were calculated using the most complete data as well as using simplified data sets. The amount of distortion was determined by comparison with the initial matrices. These processes were illustrated by stress profiles. The authors found that the loss of information was negligible if 40% of the species were deleted, and, if the individuals were not counted beyond 100.

4.4 *Dependent processes in classification and ordination space*

From a phytosociological point of view, the purpose of the analysis of various dependent processes in other spaces is in fact to find optimum sampling characteristics and data type for an adequate study and description of vegetation. Then, the main question is to what extent are the classifications and ordinations of phytosociological entities influenced by different primary processes? Past attempts to find an answer are not by far commensurable with the importance of this problem.

4.4.1 The effect of sampling

The first study on the influence of plot size upon hierarchical classifications is undoubtedly due to KERSHAW (1961) who performed normal association analysis based on two quadrat sizes. He found that three species were constant of the first four divisive species. The lower divisions were rather unstable. NOY-MEIR et al. (1970) made a broad survey of a desert community and, regarding association analysis results, similar conclusions were drawn: the divisions at lower levels were much more affected by quadrat size than at the highest level. Subdivisions at the smaller quadrat sizes were usually more meaningful than corresponding ones at larger sizes.

These two reports might suggest that plot size is not necessarily crucial in finding plant assemblage types, since the lower level changes are not significant. However, it is not known *a priori* that plot size would not considerably influence the results. It seems then that the use of various plot sizes is the most desirable sampling strategy for classifications. If the final results suggest that the group structure is stable within a broad range of plot sizes, it may be concluded that the analytical characteristic area (ACA, cf. 3.1.1) is not a fixed size, but an interval and its exact determination is unnecessary. In this case the sites consistently occurring in the same groups can be selected to find a reliable basis for the description of the types. In addition, the quadrats with uncertain group membership give useful information whether the discontinuities in the sample are strong or weak.

If considerable differences occur between classifications based on various plot sizes, the determination of the optimum cannot be avoided. Since there is a good reason to presume that the optima differ with community types, the only theoretically acceptable method would be an iterative strategy. A possible example is as follows:

1. Determine the analytical characteristic area for the whole sample, using the appropriate function f_a .
2. On the basis of the classification of quadrats equal to ACA in size, select n groups as candidates for types.
3. Find the characteristic area, CA, within each class employing function f_c . The analysis ends if
 - a) $ACA = CA$ for all i , that is the plot size corresponds to all characteristic areas, or
 - b) $ACA = CA$ for $n - 1$ types, so that a new iteration step is unnecessary. Otherwise go to the next step.
4. Now it may be assumed that the probability of misclassification is the lowest for a group whose characteristic area is best approximated by ACA. The composition of this group is considered final, and all of its quadrats are removed from the sample. Then, the ACA for the remainder is determined. The analysis returns to step 2, if the new ACA differs from the previous one. However, if there is no change in the size of ACA, other groups have to be selected and removed until the new ACA becomes different or all classes are established.

No doubt that many conditions have to be fulfilled in such an analysis, and therefore the above algorithm represents an experiment rather than a practical strategy. The range of plot sizes should be large enough to capture all characteristic areas and the size increment should be small enough to allow for the discrimination between them. Further requirement is the compatibility of functions f_a and f_c . Preferably, these functions should give minimum curves for CA (to ensure within group homogeneity) and a maximum curve for ACA (to facilitate type recognition) in terms of sample heterogeneity. Many techniques suggested to determine characteristic areas do not appear satisfactory for this purpose.

MATTHEWS (1979) substantially contributed to the problem of plot size/classification relationships. Classifications of quadrats based on four different sizes were evaluated by multiple discriminant analysis at four hierarchic levels (2-, 4-, 8- and 16-groups). There was

a good agreement between results using different quadrat sizes at the 2- and 4-cluster levels although the reduction in plot size increased the overlap between the confidence circles of groups. At levels of higher resolution (8- and 16-groups) the tendency to an increase in overlap with decreasing plot size was more marked. In general, the largest plot size gave the clearest and most interpretable result. This study implicitly suggests that at optimum plot size the discrimination among groups is the strongest. By measuring the overall discrimination, however, only an overall optimum is found, but it may be more adequate to measure the degree of overlapping separately for each type. In this way the plot size can be selected at which the given type is most clearly distinguished from the others.

When the objective of clustering is not typification but the elucidation of small-scale pattern, quadrat sampling may prove less useful as demonstrated by WILLIAMS et al. (1969). An agglomerative method, namely a type of information analysis, was applied to a pattern study in a rain-forest vegetation. Data were obtained at three quadrat sizes and with plotless sampling using four different values of neighborhood radius. Classifications of quadrats were hardly interpretable at any size, while the multiple nearest neighbor method produced meaningful results.

The problem of appropriate quadrat size in ordination studies has to be treated separately. At first sight ordinations seem less sensitive to small changes in the data than classifications. Whereas the increase of plot size may result in negligible rearrangements of points in ordination space, the same change may cause the reallocation of sample sites from one group to another. Of course, this statement has to be validated by actual studies.

Ordination of species at different plot sizes was studied by NOY-MEIR and ANDERSON (1971) in a complex manner. Variance-covariance matrices were calculated for each plot size. The matrices were added and then the combined matrix was subjected to principal component analysis (PCA) to obtain an integrated ordination of species. The eigenvalues of the combined matrix were partitioned into contributions from plot sizes. These values were plotted against plot size similarly to pattern analysis. The characteristic area of the community stand was then defined as the plot size where the most striking peak occurred. The method, known as "multiple pattern analysis", was also used by WALKER et al. (1973) and WILLIAMS et al. (1978).

FEKETE and SZÖCS (1974) used a novel approach. The 12 oak-forest species involved in the study were analyzed by PCA using different resemblance coefficients and four plot sizes. Subsequently, the species were classified by sum of squares agglomeration based on the component scores. It was demonstrated that plot size has a remarkable influence on interspecific associations. This effect was most substantial when the correlation coefficient was used.

Species interrelationships were also analyzed by MATTHEWS (1978) who used principal component analysis and multidimensional scaling (MDS) at four plot sizes. The two-dimensional arrangement of species obtained with MDS served as a basis for plexus construction. The ordinations by PCA and MDS were in general agreement, although considerable differences occurred regarding the detailed structure. Results based on the smallest plot size were the most different from the others. GOFF and MITCHELL (1975) compared species ordinations obtained at extremely different levels represented by plot and stand data. The results were fairly similar, although the degree of similarity was influenced by the type of data used.

Some comments were already made on the influence of plot shape, arrangement and sample size in sections 3.1.2-4. This problem is not discussed here again, since no quantitative analyses have been reported on this subject.

4.4.2 *The effect of processes in data space*

The dependence of ordinations upon the number of species was investigated by AUSTIN and GREIG-SMITH (1968). They found that the less abundant species contributed very little to the ordination of rain-forest stands. It is generally thought that the removal of rare species

has no practical influence on classifications (e.g., species removal prior to association-analysis is a common practice). However, this problem should be exhaustively explored using species ranking followed by successive deletion of less important species.

GOFF (1975) compared species ordinations resulting from different numbers of species. The successive changes of the initial 10-species configuration were differing with the resemblance function used. In the case of the correlation coefficient, the change was most substantial when the next five most common species were included, additional species had relatively smaller influence. By contrast, when COLE's index was used, the addition of the last six species caused the greatest change.

The use of higher taxonomic ranks in phytosociological ordination was proposed by VAN DER MAAREL (1972). He performed principal component analysis on 51 plant communities on the genus, family and order levels. Synecologically interpretable ordinations resulted, but only ordination at the order level was published and discussed. DALE and CLIFFORD (1976) employed divisive information analysis to classify a small set of data at five taxonomic levels. All classifications were similar, but similarity to the species-based classification declined as taxonomic ranks increased. It was recognized that the two highest levels (subfamily, family) were not meaningful. On the contrary, in a large-scale investigation, DEL MORAL and DENTON (1977) found the family-based classification and ordination well-interpretable. In summary it appears that the larger the scale of a given survey the higher are the taxonomic ranks appropriate for analysis.

The influence of data standardization and transformation upon the results has been discussed by several authors (see ORLÓCI 1978a, for review). However, very few studies were devoted to spatial changes whose series may be regarded as spatial processes. In fact, only two papers were found relevant, one by JENSÉN (1978) and another by VAN DER MAAREL (1979b). They both analyzed sets of phytosociological data using different transformations generated by a general exponential function derived by VAN DER MAAREL (1979a, b). Both authors have drawn the same conclusion that intermediate transformations gave always the most "acceptable" results.

4.4.3 *The number of clusters*

A fundamental question in the cluster analysis of vegetational entities is whether there exists a group structure in the data, and if so, what is the number of distinct groups? This topic will not be discussed here in detail, only a few relevant points will be made. In hierarchical clustering the cut-off level is to be determined below which the subdivision of clusters reflects no real group structure. The process of the analysis may be arrested at a termination point specified by some external criterion (stopping rule) in divisive clustering. However, in the agglomerative or the non-hierarchical methods stopping rules do not apply and the only strategy is the evaluation of a classification process represented by an increase in the number of clusters.

One possibility to test the goodness of classification is offered by discriminant analysis as shown by MATTHEWS (1979). Despite the limitations, i.e., the equality of variance-covariance matrices have to be assumed and the probability level or the dimensionality of the discriminant space are arbitrary, the method is considered robust enough to give useful information about the existence and number of clusters. A simpler alternative applicable to a series of partitions is to examine some ratio of within-group and between-group variances (e.g., RATKOWSKY and LANCE 1978, RATLIFF and PIEPER 1981) or average similarities (POPMA et al. 1983).

4.4.4 *The effect of successive algorithmic changes*

A classification process generated by successive alterations of the algorithm was explored by LANCE and WILLIAMS (1967) in connection with their flexible strategy. An "ecological data set" was classified using six values of β (3.5.1) falling within the interval $[-1, 1]$. Although the authors were unable to define rigorously the value of β , WILLIAMS (1976) in his summary stated that $\beta = -.25$ seems to meet most users' requirements.

The effect of different values for parameters α and β (3.5.3) in parametric mapping was investigated by NOY-MEIR (1974) using several sets of simulated and actual data. He found that $\alpha = 4$ and $\beta = 8$ gave the best results. However, some computational problems occurred when these values were applied to large data sets. Therefore the use of this method for the analysis of realistic data depends on the further improvement of the computer programs

5. Concluding remarks

In Vegetation Science, there are no universally applicable methods of sampling or data analysis. Although the results are not necessarily influenced much by the surveyor's decision in selecting plot characteristics, data type, resemblance function, and multivariate technique, an uncertainty is always present: one cannot predict whether the subjective choices will have significant impact on the conclusions. The present work emphasizes that the problem can be alleviated by studying the processes associated with the different stages of vegetation analysis.

Extending JUHÁSZ-NAGY's (1967, 1976) concept of the dynamism in interspecific associations, a process is defined as a series of changes in real or abstract space encountered by the survey. The different spaces themselves also constitute a series, permitting the distinction of two basic types of processes. Once the primary processes are defined in a given space, the dependent processes are automatically generated in the subsequent spaces. The most frequently used primary processes include changes in plot size and the different data manipulations. The dependent processes are usually represented by a series of classifications, ordinations, resemblance matrices or measurements of a derived variable. Further distinction can be made among primary processes according to the number of controlling parameters simultaneously changed. Accordingly the process is either elementary or complex. It is stressed that complex processes should not be studied until the effect of single controlling parameters is sufficiently known.

In the real space only primary processes can be defined. Most of these are related to one of the four sample plot characteristics; but processes can also be defined in plotless sampling. It was shown that in addition to changes of plot size and shape, sample size, plot spacing and orientation, the non-preferential designs of plot arrangement also constitute a process. The application of processes in real space may help the surveyor to optimize sampling

design, to recognize trends and pattern, and to obtain information about the stability in the results of multivariate analyses.

Manipulations of the data, with the exception of some standardization techniques, can generate primary processes in the data space. The reduction of the dimensionality of data space and the simplification of data type may be used in a pilot study to reduce sampling effort and cost in the main survey. Transformation processes are useful in examining the influence of data type on classifications and ordinations. No sharp distinction can be made between the change of the number of points in data space and the change of sample size.

The primary processes in the space of derived variables and in the resemblance space are of minor importance. In the ordination and classification spaces, processes can be defined if one or more algorithmic parameters need to be specified by the investigator. Non-hierarchical clustering strategies, and some scaling and hierarchical clustering techniques, permit specification of parameters.

It is pointed out that the different spaces are not equally important in the study of dependent processes. However, this fact alone does not account for the unbalanced treatment of the topic in the literature. Most surveys deal with the change of derived variables when affected by sampling, whereas classification and ordination processes received much less attention. One goal of the review embedded in the discussion of dependent processes was to show many unexplored possibilities.

The objective of generating primary processes is to bring about dependent processes that may be informative for various purposes. The diversity of possibilities is most striking in the case of derived variables: the precision and accuracy of statistical estimates, different vegetation parameters and the detection of pattern are all influenced by primary processes in the real space. In fact, Vegetation Science often relies on methods whose principal strategy involves a spatial process in the space of derived variables (pattern analysis, supraindividual diversity studies, etc.).

In the resemblance space, the change of interspecific associations is the most exhaustively examined process. The resemblance of quadrats, especially as affected by sampling, seems just as essential, but much less investigated.

Among the final results of vegetation analysis, the classifications and ordinations are most influenced by the surveyor's decisions. Therefore, in theory, the analysis of ordination and classification processes should constitute the basis of most vegetation surveys. Of course, it would be very expensive and cannot become an everyday practice, but some case studies can be performed to examine some principal questions. However, such a survey is conditioned by the availability of appropriate techniques for the evaluation of results.

ACKNOWLEDGEMENTS

This paper is based on the first two chapters of my Ph.D. thesis prepared at the Department of Plant Sciences, University of Western Ontario, London, Canada. I am grateful to my supervisor, Dr. L. ORLÓCI, for his guidance and encouragement throughout my research. I wish to express my sincerest thanks to Dr. R. C. JANCEY and J. B. PHIPPS (UWO), Dr. M. B. DALE (CSIRO, St. Lucia, Australia) and Dr. R. VAN HULST (Bishop's University, Lennoxville, Canada) for their comments and helpful suggestions. I have greatly benefited from discussions with Dr. P. JUHÁSZ-NAGY (ELTE, Budapest), Dr. Z. SZÓCS (Research Institute for Botany, Vácraót, Hungary) and fellow graduate students at the UWO. Many thanks are due to the Department of Plant Sciences, UWO, for the research facilities made available to me. The financial support in the form of a graduate scholarship from the Province of Ontario and the National Research Council of Canada grant given to Dr. ORLÓCI is gratefully acknowledged.

REFERENCES

- Agnew, A. D. Q. (1961): The ecology of *Juncus effusus* L. in North Wales. *J. Ecol.* **49**: 83–102.
- Austin, M. P. (1968): Pattern in a *Zerna erecta* dominated community. *J. Ecol.* **56**: 197–218.
- Austin, M. P., Greig-Smith, P. (1968): The application of quantitative methods to vegetation survey. *J. Ecol.* **56**: 827–844.
- Bartlett, M. S. (1975): *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London.
- Baum, J. D. (1964): *Elements of Point Set Topology*. Prentice Hall, Englewood Cliffs, N. J., USA.
- Blackman, G. E. (1935): A study by statistical methods of the distribution of species in grassland associations. *Ann. Bot.* **49**: 749–777.
- Bormann, F. H. (1953): The statistical efficiency of sample plot size and shape in forest ecology. *Ecology* **34**: 474–487.
- Bourdeau, P. F. (1953): A test of random versus systematic ecological sampling. *Ecology* **34**: 499–512.
- Braun-Blanquet, J. (1951): *Pflanzensoziologie. Grundzüge der Vegetationskunde*, 2nd ed., Springer, Wien.
- Bray, J. R. (1956): A study of mutual occurrence of plant species. *Ecology* **37**: 21–28.
- Cain, S. A., Castro, G. M. de O. (1959): *Manual of Vegetation Analysis*. Harper, New York.
- Chessel, D. (1981): The spatial autocorrelation matrix. *Vegetatio* **46**: 177–180.
- Clapham, A. R. (1932): The form of the observational unit in quantitative ecology. *J. Ecol.* **20**: 192–197.
- Clark, P. J., Evans, F. C. (1954): On some aspects of spatial pattern in biological populations. *Science* **121**: 397–398.
- Cliff, A. D., Ord, J. K. (1981): *Spatial Processes. Models and Applications*. Pion, London.
- Cochran, W. G. (1963): *Sampling Techniques*. Wiley, New York.
- Cormack, R. M. (1971): A review of classification. *J. R. Stat. Soc. ser. A* **134**: 321–367.
- Cormack, R. M., Ord, J. K. (eds) (1979): *Spatial and Temporal Analysis in Ecology*. International Co-operative Publ. House, Fairland, Maryland, USA.
- Cottam, G. (1947): A point method for making rapid surveys of woodlands. *Bull. ecol. Soc. Amer.* **28**: 60.
- Cottam, G., Curtis, J. T. (1949): A method for making rapid surveys of woodlands by means of pairs of randomly selected trees. *Ecology* **30**: 101–104.
- Cottam, G., Curtis, J. T., Hale, B. W. (1953): Some sampling characteristics of a population of randomly dispersed individuals. *Ecology* **34**: 741–757.
- Crawford, R. M. M., Wishart, D. W. (1968): A rapid classification and ordination method and its application to vegetation mapping. *J. Ecol.* **56**: 385–404.
- Crovello, T. J. (1968): The effect of change of number of OTUs in a numerical taxonomic study. *Brittonia* **20**: 346–367.
- Curtis, J. T., McIntosh, R. P. (1950): The interrelations of certain analytic and synthetic phytosociological characters. *Ecology* **31**: 434–455.
- Dale, M. B. (1977): Planning an adaptive numerical classification. *Vegetatio* **35**: 131–136.
- Dale, M. B., Clifford, H. T. (1976): On the effectiveness of higher taxonomic ranks for vegetation analysis. *Austr. J. Ecol.* **1**: 37–62.
- Dale, M. B., Williams, W. T. (1978): A new method of species reduction for ecological data. *Aust. J. Ecol.* **3**: 1–5.

- Dévai, I., Horváth, K., Juhász-Nagy, P. (1971): Some problems of model-building in syn-biology. Part I. Spatial diversity process of the binary type in a simple situation. *Annal. Univ. Sci. Budapest, sect. biol.* **13**: 19–32.
- Dietvorst, P., van der Maarel, E., van der Putten, H. (1982): A new approach to the minimal area of a plant community. *Vegetatio* **50**: 77–91.
- Diggle, P. J. (1979): Statistical methods for spatial point patterns in ecology. In: Cormack–Ord (op. cit.), pp. 95–150.
- Du Rietz, G. E. (1921): Zur methodologischen Grundlagen der modernen Pflanzensoziologie. *Akad. Aft., Uppsala*.
- Ebert, T. A., McMaster, G. S. (1981): Regular pattern of desert shrubs: A sampling artefact? *J. Ecol.* **69**: 559–564.
- Estabrook, G. F. (1966): A mathematical model in graph theory for biological classification. *J. theoret. Biol.* **12**: 297–310.
- Evans, F. C. (1952): The influence of size of quadrat on the distributional patterns of plant populations. *Contr. Lab. Vertebr. Biol. Univ. Mich.* **54**: 1–15.
- Everitt, B. S. (1979): Unresolved problems in cluster analysis. *Biometrics* **35**: 169–181.
- Falinski, J. B. (ed.) (1977): *Vegetation Dynamics*. Phytocoenosis 7.
- Fekete, G., Kovács, M. (1978): The space dynamism of species and life form diversity in two rocky grassland communities. *Acta Biol. Debrecina* **15**: 7–21.
- Fekete, G., Szócs, Z. (1974): Studies on interspecific association processes in space. *Acta Bot. Acad. Sci. Hung.* **20**: 227–241.
- Fekete, G., Szujkó-Lacza, J. (1973): Interspecific correlation of plants in oakwood at increasing block sizes. *Acta Biol. Acad. Sci. Hung.* **24**: 31–42.
- Feoli, E., Lagonegro, M., Orlóci, L. (1982): *Information Analysis in Vegetation Research*. Prepublication edition, London–Trieste.
- Fewster, P., Orlóci, L. (1978): Stereograms to aid group recognition and trend identification in vegetation data. *Can. J. Bot.* **56**: 162–165.
- Finney, D. J. (1948): Random and systematic sampling in timber surveys. *Forestry* **22**: 64–99.
- Finney, D. J. (1950): An example of periodic variation in forest sampling. *Forestry* **23**: 96–111.
- Forgy, E. W. (1965): Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* **21**: 768–769.
- Fowler, G. W., Davis, C. F. (1979): Sampling natural resource populations: Mutually exclusive fixed-area sampling units. *Resource Inventory Notes (Denver) USDI BLM-23*: 1–7.
- Fowler, H. W., Fowler, F. G. (1956): *The Concise Oxford Dictionary*. Oxford Univ. Press.
- Freese, F. (1961): Relation of plot size to variability: an approximation. *J. Forestry* **58**: 679.
- Freund, J. E. (1967): *Modern Elementary Statistics*. 3rd edition. Prentice-Hall, Englewood Cliffs, N. J.
- Gauch, H. G. (1980): Rapid initial clustering of large data sets. *Vegetatio* **42**: 103–111.
- Gauch, H. G. (1982): *Multivariate Analysis of Community Data*. Cambridge Univ. Press.
- Gittins, R. (1969): The application of ordination techniques. In: I. H. Rorison (ed.), *Ecological Aspects of the Mineral Nutrition of Plants*. pp. 37–66. Symp. No. 9, Blackwell, Oxford.
- Gloaguen, J. C., Gautier, N. (1981): Pattern development of the vegetation during colonization of a burnt heathland in Brittany (France). *Vegetatio* **46**: 167–176.
- Goff, F. G. (1975): Comparison of species ordinations resulting from alternative indices of interspecific association and different numbers of included species. *Vegetatio* **31**: 1–14.
- Goff, F. G., Mitchell, R. (1975): A comparison of species ordination results from plot and stand data. *Vegetatio* **31**: 15–22.
- Goodall, D. W. (1954): Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Aust. J. Bot.* **2**: 304–324.
- Goodall, D. W. (1961): Objective methods for the classification of vegetation. IV. Pattern and minimal area. *Aust. J. Bot.* **9**: 162–196.
- Goodall, D. W. (1963): Pattern analysis and minimal area — some further comments. *J. Ecol.* **51**: 705–710.
- Goodall, D. W. (1970): Statistical plant ecology. *Ann. Rev. Ecol. Syst.* **1**: 99–124.
- Goodall, D. W. (1974): A new method for the analysis of spatial pattern by random pairing of quadrats. *Vegetatio* **29**: 135–146.
- Gounot, M. (1969): *Methodes d'Étude Quantitative de la Vegetation*. Paris.
- Gounot, M., Calléja, M. (1962): Coefficient de communauté, homogénéité et aire minimale. *Bull. Serv. Carte Phytogéogr., Ser. B.* **7**: 181–200.
- Grassle, J. P., Grassle, J. F. (1976): Sibling species in the marine pollution indicator *Capitella* (Polychaeta). *Science* **192**: 567–569.
- Green, R. H. (1979): *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, New York.

- 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.
- Greig-Smith, P. (1979): Pattern in vegetation. *J. Ecol.* **67**: 755–779.
- Grime, J. P. (1979): *Plant Strategies and Vegetation Processes*. Wiley, New York.
- Hall, J. B., Okali, D. U. U. (1979): A structural and floristic analysis of woody fallow vegetation near Ibadan, Nigeria. *J. Ecol.* **67**: 321–346.
- Hasel, A. A. (1938): Sampling errors in timber surveys. *J. Agr. Res.* **57**: 713–736.
- Heyer, W. R., Berven, K. A. (1973): Species diversities of herpetofaunal samples from similar microhabitats at two tropical sites. *Ecology* **54**: 642–645.
- Hill, M. O. (1973a): The intensity of spatial pattern in plant communities. *J. Ecol.* **61**: 225–235.
- Hill, M. O. (1973b): Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**: 427–432.
- Hopkins, B. (1955): The species-area relations of plant communities. *J. Ecol.* **43**: 409–426.
- Jancey, R. C. (1966): Multidimensional group analysis. *Austr. J. Bot.* **14**: 127–130.
- Jancey, R. C. (1974): Algorithm for detection of discontinuities in data sets. *Vegetatio* **29**: 131–133.
- Jancey, R. C. (1979): Species weighting: A heuristic approach. In: Orlóci et al. (op. cit.), pp. 87–100.
- Jancey, R. C. (1980): The minimisation of random events in the search for group structure. *Vegetatio* **42**: 99–101.
- Janssen, J. G. M. (1975): A simple clustering procedure for preliminary classification of very large sets of phytosociological results. *Vegetatio* **30**: 67–71.
- Jardine, N., Sibson, R. (1971): *Mathematical Taxonomy*. Wiley, New York.
- Jensén, S. (1978): Influences of transformation of cover values on classification and ordination of lake vegetation. *Vegetatio* **37**: 19–31.
- Juhász-Nagy, P. (1967): On some "characteristic areas" of plant community stands. *Proc. Colloq. Inform. 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., Dévai, I., Horváth, K. (1973): Some problems of model-building in synbiology. Part 2. Associatum process in a simple situation. *Annal. Univ. Sci. Budapest, sect. biol.* **15**: 39–51.
- Juhász-Nagy, P., Podani, J. (1983): Information theory methods for the study of spatial processes and succession. *Vegetatio* **51**: 129–140.
- Justesen, S. H. (1932): Influence of size and shape of plots on the precision of field experiments with potatoes. *J. Agric. Sci.* **22**: 366–372.
- Kaesler, R. L., Cairns, J., Crossman, J. S. (1974): Redundancy in data from stream surveys. *Water Res.* **8**: 637–642.
- Kalamakar, R. J. (1932): Experimental error and the field plot technique with potatoes. *J. Agric. Sci.* **22**: 373–385.
- Kenkel, N. (1981): Sampling theory in quantitative plant ecology. Manuscript, Department of Plant Sciences, UWO, London, Ontario.
- Kershaw, K. A. (1957): The use of cover and frequency in the detection of pattern in plant communities. *Ecology* **38**: 291–299.
- Kershaw, K. A. (1961): Association and co-variance analysis of plant communities. *J. Ecol.* **49**: 643–654.
- Kershaw, K. A. (1973): *Quantitative and Dynamic Plant Ecology*. Edward Arnold, London.
- Kilburn, P. D. (1966): Analysis of the species-area relation. *Ecology* **47**: 831–843.
- Kobayashi, S. (1981): Diversity indices: Relations to sample size and spatial distribution. *Jap. J. Ecol.* **31**: 231–236.
- Kruskal, J. B. (1964): Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**: 115–129.
- 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.
- Lambert, J. M., Dale, M. B. (1964): The use of statistics in phytosociology. *Adv. Ecol. Res.* **2**: 59–99.
- Lance, G. N., Williams, W. T. (1967): A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer J.* **9**: 373–380.
- Ludwig, J. A. (1979): A test of different quadrat-variance methods for the analysis of spatial pattern. In: Cormack and Ord (op. cit.), pp. 289–304.

- Ludwig, J. A., Goodall, D. W. (1978): A comparison of paired- with blocked-quadrat variance methods for the analysis of spatial pattern. *Vegetatio* **38**: 49–59.
- Maarel, E. van der (1972): Ordination of plant communities on the basis of their plant genus, family and order relationships. In: E. van der Maarel and R. Tüxen (eds), *Grundfragen und Methoden in der Pflanzensoziologie*. pp. 183–206. Junk, The Hague.
- Maarel, E. van der (1979a): Multivariate methods in phytosociology, with reference to the Netherlands. In: Wergler, M. J. A. (ed.), *The Study of Vegetation*. pp. 161–225. Junk, The Hague.
- Maarel, E. van der (1979b): Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* **39**: 97–114.
- Maarel, E. van der (ed.) (1980): Advances in vegetation science: succession. *Vegetatio* **43**: 1–151.
- MacNaughton-Smith, P., Williams, W. T., Dale, M. B., Mockett, L. G. (1964): Dissimilarity analysis: a new technique of hierarchical sub-division. *Nature* **202**: 1034–1035.
- MacQueen, J. B. (1967): Some methods for classification and analysis of multivariate observations. *Proc 5th Berkeley Symposium on Math. Stat. and Prob.* **1**: 281–297.
- Matthews, J. A. (1978): An application of non-metric multidimensional scaling to the construction of an improved species plexus. *J. Ecol.* **66**: 157–173.
- Matthews, J. A. (1979): A study of the variability of some successional and climax plant assemblage-types using multiple discriminant analysis. *J. Ecol.* **67**: 255–271.
- McIntosh, R. P. (1973): Matrix and plexus techniques. In: Whittaker, R. H. (op. cit.), pp. 157–191.
- Molnár, E. N., Nosek, J. N. (1979): Spatial processes in a grassland community. I. Number of species and individuals, cover and biomass at the community level. *Acta Bot. Acad. Sci. Hung.* **25**: 339–348.
- Molnár, E. N., Nosek, J. N. (1980): Spatial processes in a grassland community. II. Number of individuals, cover and biomass at population level. *Acta Bot. Acad. Sci. Hung.* **26**: 429–442.
- Moral, R. del, Denton, M. F. (1977): Analysis and classification of vegetation based on family composition. *Vegetatio* **34**: 155–165.
- Moravec, J. (1973): The determination of the minimal area of phytocenoses. *Folia Geobot. Phytotax.* **8**: 23–47.
- Morisita, M. (1954): Estimation of population density by spacing method. *Mem. Fac. Sci. Kyushu Univ. Series E. (Biol.)* **1**: 187–197.
- Morton, A. J. (1974): Ecological studies of a fixed dune grassland at Newborough Warren, Anglesey. I–II. *J. Ecol.* **62**: 253–260, 261–278.
- Myers, E., Chapman, V. J. (1953): Statistical analysis applied to a vegetation type in New Zealand. *Ecology* **34**: 175–185.
- 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.
- Noy-Meir, I. (1971): Multivariate analysis of desert vegetation. II. Qualitative/quantitative partition of heterogeneity. *Israel J. Bot.* **20**: 203–213.
- Noy-Meir, I. (1974): Catenation: quantitative methods for the definition of coenoclines. *Vegetatio* **29**: 89–99.
- Noy-Meir, I., Tadmor, N. H., Orshan, G. (1970): Multivariate analysis of desert vegetation. I. Association analysis at various quadrat sizes. *Israel J. Bot.* **19**: 561–591.
- Noy-Meir, I., Anderson, D. J. (1971): Multiple pattern analysis, or multiscale ordination: towards a vegetation hologram? In: G. P. Patil, E. C. Pielou, W. E. Waters (eds): *Many Species Populations, Ecosystems, and Systems Analysis. Stat. Ecol.* **3**: 207–225. Penn. State Univ. Press.
- 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. Manage.* **37**: 217–222.
- Orlóci, L. (1967): An agglomerative method for classification of plant communities. *J. Ecol.* **55**: 193–205.
- Orlóci, L. (1969): Information theory models for hierarchic and non-hierarchic classifications. In: A. J. Cole (ed.): *Numerical Taxonomy*. pp. 148–164. Academic Press, London.
- Orlóci, L. (1971): An information theory model for pattern analysis. *J. Ecol.* **59**: 343–349.
- Orlóci, L. (1973): Ranking characters by a dispersion criterion. *Nature* **244**: 371–373.
- Orlóci, L. (1976a): Ranking species by an information criterion. *J. Ecol.* **64**: 417–419.
- Orlóci, L. (1976b): TRGRPS — An iterative algorithm for group recognition with an example from *Spartinetea*. *Vegetatio* **32**: 117–120.
- Orlóci, L. (1978a): *Multivariate Analysis in Vegetation Research*. Junk, The Hague.

- Orlóci, L. (1978b): Ranking species based on the components of equivocation information. *Vegetatio* **37**: 123-125.
- Orlóci, L. (1980): An algorithm for predictive ordination. *Vegetatio* **42**: 23-25.
- Orlóci, L., Mukkattu, M. M. (1973): The effect of species number and type of data on the resemblance structure of a phytosociological collection. *J. Ecol.* **61**: 37-46.
- Orlóci, L., Rao, C. R., Stiteler, W. M. (eds) (1979): *Multivariate Methods in Ecological Work*. International Co-operative Publ. House, Fairland, Maryland, USA.
- Peckham, J. F., Stewart, G. (1940): Sagebrush-grass range sampling studies: size and structure of sampling unit. *J. Amer. Soc. Agron.* **32**: 669-682.
- Peet, R. K. (1974): The measurement of species diversity. *Ann. Rev. Ecol. Syst.* **5**: 285-307.
- Penfound, W. T. (1945): A study of phytosociological relationships by means of aggregations of colored cards. *Ecology* **26**: 38-57.
- Pielou, E. C. (1969): *An Introduction to Mathematical Ecology*. Wiley, New York.
- Pielou, E. C. (1975): *Ecological Diversity*. Wiley, New York.
- Pielou, E. C. (1977): *Mathematical Ecology*. Wiley, New York.
- Podani, J. (1982): Spatial processes in the analysis of vegetation. Ph.D. Thesis. Department of Plant Sciences, University of Western Ontario, London, Canada.
- Poole, R. W. (1974): *An Introduction to Quantitative Ecology*. McGraw-Hill, New York.
- Popma, J., Mucina, L., Tongeren, O. van, Maarel, E. van der (1983): On the determination of optimal levels in phytosociological classification. *Vegetatio* **52**: 65-75.
- Ram Babu, Agarwal, M. C., Puri, D. N. (1981): Size and shape of plots and blocks for field experiments in natural grasslands in Agra ravines. *Indian J. Agric. Sci.* **51**: 271-275.
- Ratkowsky, D. A., Lance, G. N. (1978): A criterion for determining the number of groups in a classification. *Austr. Computer J.* **10**: 115-117.
- Ratcliff, D., Pieper, R. D. (1981): Deciding final clusters: An approach using intra- and inter-cluster distances. *Vegetatio* **43**: 83-86.
- Raunkiaer, C. (1934): *The Life-forms of Plants and Statistical Plant Geography*. Oxford Univ. Press.
- Rényi, A. (1961): On measures of entropy and information. *Proc. 4th Berkeley Symposium on Math. Stat. and Prob.* **1**: 547-561.
- Rice, E. L. (1967): A statistical method for determining quadrat size and adequacy of sampling. *Ecology* **48**: 1047-1049.
- Ripley, B. D. (1978): Spectral analysis and the analysis of pattern in plant communities. *J. Ecol.* **66**: 965-981.
- Rogers, A. (1974): *Statistical Analysis of Spatial Dispersion*. The Quadrat Method. Pion, London.
- Rogers, R. S. (1980): Plot distribution in the understory of a Beech-Maple forest: a comparison of two contrasting sampling designs. *Bull. Torrey Bot. Club.* **107**: 530-538.
- Rohlf, F. J., Sokal, R. R. (1981): Comparing numerical taxonomic studies. *Syst. Zool.* **30**: 459-490.
- Sampford, M. R. (1962): *An Introduction to Sampling Theory*. Oliver & Boyd, Edinburgh.
- Seal, H. L. (1964): *Multivariate Statistical Analysis for Biologists*. Methuen, London.
- Shepard, R. N., Carroll, J. D. (1966): Parametric representation of nonlinear data structures. In: Krishnaiah, P. R. (ed.), *Multivariate Analysis*, pp. 561-592. Academic Press, London.
- Shugart, H. H. (ed.) (1978): *Time Series and Ecological Processes*. SIAM, Philadelphia.
- Sneath, P. H. A. (1966): A method for curve seeking from scattered points. *Computer J.* **8**: 383-391.
- Sneath, P. H. A., Sokal, R. R. (1973): *Numerical Taxonomy*. Freeman, San Francisco.
- Sokal, R. R., Rohlf, F. J. (1962): The comparison of dendrograms by objective methods. *Taxon* **11**: 33-40.
- Sokal, R. R., Rohlf, F. J. (1973): *Introduction to Biostatistics*. Freeman, San Francisco.
- Szőcs, Z. (1979): New computer-oriented methods for the study of natural and simulated vegetation structure. In: Orlóci et al. (op. cit.), pp. 301-308.
- Taaffe, K. E. (1979): Computing optimum plot size for wildland inventories. *Resource Inventory Notes (Denver) USDI BLM-23*: 8-15.
- Thomas, H. L., Abou-El-Fittou, H. A. (1968): Optimum plot size and number of replications for estimating forage yield and moisture percentage. *Agronomy J.* **60**: 549-550.
- Thompson, H. R. (1956): Distribution of distance to nth neighbor in a population of randomly distributed individuals. *Ecology* **37**: 391-394.
- Trass, H., Malmer, N. (1973): North European approaches to classification. In: Whittaker, R. H. (op. cit.), pp. 531-574.

- Tüxen, R. (1970): Bibliographie zum Problem des Minimum-Areals und der Art-Areal Kurve. *Excerpta Bot., Sect. B.* **10**: 291–314.
- Usher, M. B. (1969): The relation between mean square and block size in the analysis of similar patterns. *J. Ecol.* **57**: 505–514.
- Usher, M. B. (1975): Analysis of pattern in real and artificial plant populations. *J. Ecol.* **63**: 569–586.
- van Dyne, G. M., Vogel, W. G., Fisser, H. G. (1963): Influence of small plot size and shape on range herbage production estimates. *Ecology* **44**: 746–759.
- van Groenewoud, H., Ihm, P. (1974): Cluster analysis based on graph theory. *Vegetatio* **29**: 115–120.
- Walker, J., Noy-Meir, I., Anderson, D. J., Moore, R. M. (1972): Multiple pattern analysis of a woodland in South Central Queensland. *Austr. J. Bot.* **20**: 105–118.
- Watt, A. S. (1947): Pattern and process in the plant community. *J. Ecol.* **35**: 1–22.
- Weaver, J. E., Clements, F. E. (1938): *Plant Ecology*. 2. ed. New York.
- Webb, L. J., Tracey, J. G., Williams, W. T., Lance, G. N. (1967): Studies in the numerical analysis of complex rain-forest communities. II. The problem of species-sampling. *J. Ecol.* **55**: 525–538.
- West, O. (1937): An investigation of the methods of botanical analysis of pasture. *S. Afr. J. Sci.* **33**: 501–559.
- Westhoff, V. (1951): An analysis of some concepts and terms in vegetation study or phytocenology. *Synthese* **8**: 194–206.
- Westhoff, V., van der Maarel, E. (1973): The Braun-Blanquet approach. In: Whittaker, R. H. (op. cit.), pp. 619–726.
- White, P. S. (1979): Pattern, process, and natural disturbance in vegetation. *Bot. Rev.* **45**: 229–299.
- Whittaker, R. H. (1967): Gradient analysis of vegetation. *Biol. Rev.* **42**: 207–264.
- Whittaker, R. H. (1973a): Direct gradient analysis: Techniques. In: Whittaker, R. H. (op. cit.), pp. 9–31.
- Whittaker, R. H. (ed.) (1973b): *Ordination and Classification of Communities*. Junk, The Hague.
- Wiegert, R. G. (1967): The selection of an optimum quadrat size for sampling the standing crop of grasses and forbs. *Ecology* **43**: 125–129.
- Wildi, O., Orłóci, L. (1980): Management and Multivariate Analysis of Vegetation Data. *Eidg. Aust. forstl. Versuchsv. Ber. v.* 215.
- Williams, D. G., Anderson, D. J., Slater, K. R. (1978): The influence of sheep on pattern and process in *Atriplex vesicaria* populations from the Riverine Plain of New South Wales. *Aust. J. Bot.* **26**: 381–392.
- Williams, W. T. (ed.) (1976): *Pattern Analysis in Agricultural Science*. Elsevier, Amsterdam.
- Williams, W. T., Dale, M. B. (1965): Fundamental problems in numerical taxonomy. *Adv. Bot. Res.* **2**: 35–68.
- 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.
- Wolda, H. (1981): Similarity indices, sample size and diversity. *Oecologia* **50**: 296–302.
- Zahl, S. (1974): Application of the S-method to the analysis of spatial pattern. *Biometrics* **30**: 513–524.
- Zeide, B. (1980): Plot size optimization. *Forest Sci.* **26**: 251–257.