

On sampling procedures in population and community ecology

N. C. Kenkel,¹ P. Juhász-Nagy² & J. Podani²

¹*Department of Botany, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada;* ²*Department of Plant Ecology and Taxonomy, L. Eötvös University, Kun Béla tér 2, Budapest H-1083, Hungary*

Accepted 20.3.1989

Keywords: Estimation, Multivariate, Plot, Sample size, Spatial pattern, Univariate

Abstract

In this paper we emphasize that sampling decisions in population and community ecology are context dependent. Thus, the selection of an appropriate sampling procedure should follow directly from considerations of the objectives of an investigation. We recognize eight sampling alternatives, which arise as a result of three basic dichotomies: parameter estimation versus pattern detection, univariate versus multivariate, and a discrete versus continuous sampling universe. These eight alternative sampling procedures are discussed as they relate to decisions regarding the required empirical sample size, the selection or arrangement of sampling units, and plot size and shape. Our results indicate that the decision-making process in sampling must be viewed as a flexible exercise, dictated not by generalized recommendations but by specific objectives: there is no panacea in ecological sampling. We also point to a number of unresolved sampling problems in ecology.

Introduction

It has long been recognized that sampling procedures play an important role in population and community studies in ecology (Greig-Smith 1983; Green 1979). Over the past few decades, however, the clarification of sampling objectives and the elucidation of sampling problems in ecology have received little attention. Indeed, ecologists have generally applied uncritically the well-established principles and procedures of classical sampling theory (Cochran 1977) to ecological problems. While realizing that many of these fundamental principles (e.g. randomization) underly all sampling decisions, it is also important to recognize that the objectives of an ecological study may differ from those considered by the classical theory. Specifically, classical sampling theory is

largely concerned with population parameter estimation, in which the sampling units are discrete, recognizable entities. Ecological investigations, by contrast, often involve pattern recognition in communities, in which the sampling unit is arbitrarily defined (e.g. a plot). Failure to fully recognize such differences has resulted in the superficial treatment of many sampling problems in ecology.

Why is there a need to clarify sampling principles and procedures in ecology? We note that in recent years considerable progress has taken place in the development of methodological procedures for analyzing multivariate data (Digby & Kempton 1987; Ludwig & Reynolds 1988) and spatial pattern (Upton & Fingleton 1985). During the same period, however, many fundamental problems in ecological sampling have been all but ignored (Greig-Smith 1983). Because sampling

represents the first step in an ecological study, sampling decisions will necessarily affect information flow and so impact upon final conclusions. It is therefore imperative that sampling decisions be made which are consistent with the objectives of a study (c.f. Sukhatma *et al.* 1984, p. 4).

Very general sampling procedures are often recommended in the ecological literature, which may create problems if the objectives of the study are at variance with the recommended procedure. As a simple example, consider the problem of determining an adequate sample size. The solution is quite straightforward given the objective of population parameter estimation (e.g. mean or total). First, a small preliminary random sample is taken from the population to estimate the sampling variance. A simple formula is then used to calculate the sample size required to obtain an estimate within specified probability limits (e.g. Green 1979, pp. 40, 126; Elliott 1983, p. 128). This technique, while useful, is largely irrelevant when the objective is pattern recognition. Furthermore, it is applicable only to the population (single species) case. Sample size determination under objectives other than population parameter estimation is not discussed in monographs on community ecology, however.

A further problem is that very general statements regarding sampling decisions are made in the ecological literature. Should such statements be taken out of context, sampling procedures which are inconsistent with stated objectives may be used. As an example, consider the following statement in a discussion of plot shape (Gauch 1982, p. 57):

‘On the whole, a rectangle that is two to four times as long as it is wide is ordinarily most accurate...’

Such a general statement invites a number of questions: How can a single sampling unit be accurate? What is meant by accuracy in this context? What does ‘ordinarily’ imply? And most importantly, under what objectives does this statement hold? As we shall see, the determination of plot shape should be dictated by the

objectives of a study. Many other examples of this problem could be cited from the literature: suffice to say that sampling decisions must be dictated by the study objectives, rather than by recommended ‘general principles’.

The objective of this review is to point out that ecological studies have very divergent objectives, and that these objectives are of fundamental importance in making intelligent and meaningful sampling decisions. We begin our discussion by introducing some basic terms and concepts, and follow this with a discussion of eight sampling alternatives. We then examine the problem of determining sample size and the selection (or spatial arrangement) of sampling units, and discuss problems in the determination of plot size and shape. Our hope is to clarify some of the existing confusion in the ecological literature regarding sampling, and to indicate problem areas which are deserving of greater attention.

Basic terms and concepts

Field biologists obtain information in one or two ways: *observational* (reconnaissance of an area, in which no data are recorded) or *analytical* (data collected, usually for further processing). Data collection necessarily involves the application of a *sampling procedure* (measurements, counts, and so forth). Unfortunately, some confusion has arisen in the ecological literature due to the misuse of well-established sampling theory terminology. We therefore begin by clarifying some terms and concepts before turning to a discussion of problems specific to sampling in population and community ecology.

In statistics, a universal set of data is referred to as a *population*, a term which is a potential source of confusion. It was originally a biological term, and has been used in this context since the 17th century (for example, by Malthus). In biological studies, therefore, the biological definition of the term should receive priority over its statistical use. We propose *sampling universe* to replace population as a statistical term. Note that in many cases there is a fairly good correspondence

between a population and a sampling universe: thus, a trait is Mendelian if and only if the sample is rightly supposed to be drawn from a Mendelian sampling universe.

A clear distinction should be made between the *empirical* and *theoretical* aspects of sampling. Theoretical sampling refers to the derivation of methods of inference which define a sample, such that specified statistical conditions (e.g. representativeness, unbiasedness, and so forth) are met. By contrast, empirical sampling refers to the actual process of data collection, and is associated with terms such as *sampling units* and *sampling design*. Sampling units are usually considered to be discrete entities whose assemblage corresponds to the sampling universe (e.g. Cochran 1977, p. 6), while sampling design refers to the selection or arrangement of these units. Empirical and theoretical aspects should be simultaneously considered in sampling surveys. Ideally a methodological iteration, in which conditions of empirical sampling are modified so as to approximate theoretical requirements, is the most admissible strategy.

The term sample size is sometimes used by ecologists to refer to the size of an individual sampling unit (Gauch 1982; Greig-Smith 1983). We prefer the statistical definition, and therefore refer to the number of sampling units in a sample as the *empirical sample size*. By contrast, the *theoretical sample size* is the number of data items in the sample. The empirical and theoretical sample sizes will therefore be equal only if a single variable is measured for each sampling unit. In multivariate studies, the multiple of the number of sampling units and variables recorded on each will provide the theoretical sample size.

Sampling theory assumes that a random process is involved at some stage of the selection of the sampling units (*probabilistic sampling*, Cochran 1977, p. 9). By contrast, the ill-defined *non-probabilistic sampling* involves selection from a restricted, accessible part of the population, or the preferential selection of what are considered to be 'typical' units. The latter practice, which is widespread among phytosociologists, takes advantage of the previous experience of the researcher (e.g.

Braun-Blanquet 1964). Note, however, that statistical inference is valid only if probabilistic sampling is undertaken. We do not consider non-probabilistic sampling in this paper, as it is well covered elsewhere (Knapp 1984).

Sampling alternatives

The sampling procedure used is ultimately determined by three dichotomous choices: study objectives (estimation or pattern recognition), number of variables (univariate or multivariate), and the nature of the sampling universe (discrete or continuous). Eight distinct sampling alternatives can therefore be recognized (Table 1). As the examples in the table indicate, the determination of a suitable sampling strategy is context-dependent: an appropriate procedure for any one of the eight alternatives may be unsuitable for another. This implies that general statements regarding sampling decisions (specifically with respect to empirical sample size, the arrangement or selection of sampling units, and the size and shape of sampling units) are of limited utility if not placed in their proper context.

Parameter estimation vs. pattern detection

In many fields of biology sampling is undertaken to obtain estimates of population parameters (e.g. mean cover or density, diversity). The estimator, a statistical function, should meet several basic requirements including unbiasedness, consistency, efficiency, sufficiency and minimum variance (Wilks 1962, p. 351). Classical sampling theory, which is concerned with both the derivation of estimators and the maximization of precision of estimates through alternative sampling designs, is appropriate for this purpose. Precision in this context refers to the minimization of the variability (sampling variance) about the estimated parameter.

In ecology, estimation is often not the sole aim of a study. Instead, the principal focus may be on the analysis of pattern: here, we define pattern in

Table 1. Dichotomous sampling alternatives in plant ecology.

Objective of the study	Number of variables	Nature of sampling universe	Examples
Parameter estimation	Univariate	Discrete Continuous	Tree height, seed production per plant Cover, density, biomass and diversity estimation by plots. Random plant methods for estimating density and cover.
	Multivariate	Discrete Continuous	Estimation of the mean vector of morphological variables. Estimation of the mean vector of cover values.
Pattern analysis	One population	Discrete	Random plant methods.
		Continuous	Quadrat-variance methods. Random points methods.
	Community	Discrete Continuous	Ordination or classification of discrete entities (islands, lakes, epiphytic assemblages, etc.) Florula diversity, resemblance, pairwise interspecific correlation. Ordination and classification of plots.

its broad sense, to include classification, ordination, and detection of the intensity and scale of spatial variation. It is clear that sampling decisions under an objective of parameter estimation differ from those taken when pattern recognition is the stated objective. As a simple example consider the problem of plot size determination. For simplicity, assume that a species X has a clumped pattern, and that n sampling units (plots) are to be taken from within the study area. To obtain a precise estimate of mean cover for the species, we would choose a plot size such that the between-plot variability is *minimized*. For this purpose, a plot size somewhat larger than the mean clump size would likely be appropriate. For pattern recognition, however, plot size should be chosen so as to *maximize* between-plot variability. Under such circumstances, a plot of the same size as the mean species clump size should be utilized.

Univariate vs. multivariate

In estimation, univariate refers to the determination of a single random variable, regardless of whether one or several species are considered. Examples include mean cover or yield (of a single species or the whole community) and textural variables such as species-individual diversity (sensu Barkman 1979). Univariate estimation problems are well covered by sampling theory

(Cochran 1977). In the multivariate situation several variables are simultaneously measured, and estimates are obtained for elements of a mean vector: here, the covariance (correlation) structure of variables must also be considered.

In pattern analysis, the term univariate refers to recognition of the scale of pattern of a single species. Studies of community pattern are multivariate (multi-species): examples include the analysis of interspecific association, ordination and classification.

Discrete vs. continuous sampling universe

The sampling universe is said to be discrete when the sampling units are natural, distinct, and recognizable entities. It is therefore of finite size N , implying a finite number of possible samples (2^N). Under these circumstances the sampling frame (a list of individuals making up the sampling universe) is definable. Examples of discrete sampling units include individual plant genets, and isolated biogeographic units such as islands or lakes.

In a continuous sampling universe there are no natural, recognizable sampling units. Instead the sampling units, which may have dimensions of zero (point-intercept), one (line-intercept), two (plot or quadrat), or three (soil core or benthic grab), are arbitrarily delineated by the investiga-

tor. Since the sampling units are arranged within a spatial continuum, the statistical population is of infinite size and a sampling frame is therefore not defined.

The nature of the sampling universe has important consequences for decision-making in sampling. Given a discrete sampling universe (the type treated in sampling theory), one need only be concerned with the manner by which individuals are chosen from the sampling frame: alternatives include random, and various restricted randomized procedures (e.g. stratification, systematic, or cluster sampling). Conversely, in a continuous sampling universe one must consider the spatial arrangement of sampling units, as well as their size and shape (for the two and three-dimensional cases).

Empirical sample size

Univariate estimation

Determination of sample size is straightforward for the discrete sampling universe case. The method is based on the relationship

$$S_{\bar{x}} = [S^2/n]^{1/2} \quad (1)$$

where n is the sample size, S^2 is the variance, and $S_{\bar{x}}$ is the standard error of the mean. The sampling universe is assumed to be infinitely large. The relationship indicates that any increase in the sample size necessarily decreases the standard error of the mean. Biological examples of empirical sample size determination can be found in Green (1979, pp. 40, 126), Elliott (1983, p. 127), Greig-Smith (1983), and De Vries (1986). While the same method can be used when the sampling universe is continuous, it must be remembered that the variance of the sample will depend on the size and shape of the sampling unit (e.g. a plot). For a clumped pattern, a plot size of the same order as the mean clump size will result in a large variance. Thus the sample size required to achieve a given level of precision will be much greater than would be required if the plot size were somewhat

larger than the mean clump size. Note that if the total area which can be sampled is fixed (as determined by some cost function), it is generally recommended that many small plots rather than a few large ones be used (Green 1979, pp. 39, 132).

For diversity measures, the variation of an estimate tends to decrease as sample size increases (Pielou 1977, p. 301). A simulation study by Wolda (1981) demonstrated that the standard deviation of diversity estimates obtained by replicate samples decreased as sample size increased. Sample size effects on evenness estimates cannot be similarly interpreted, however. Species evenness tends to decrease when an increase in sample size leads to the occurrence of new species in the sample. Evenness estimates therefore require very careful interpretation (Pielou 1977, p. 307).

Multivariate estimation

To obtain simultaneous estimates of several variables one could simply use the univariate procedure outlined above, applying it separately to each variable. This will very likely result in vastly different sample size determinations for each variable, however (e.g. Croy & Dix 1984). Furthermore, such a strategy fails to consider the covariance structure of the variable set. The method of Scagel *et al.* (1985) attempts to overcome this problem by determining the empirical sample size required to stabilize the correlation structure of multivariate data. Unfortunately, their method requires a large initial sample size.

The above arguments hold for both discrete and continuous sampling universes. In the latter case, the precision of estimates will be influenced by the size and shape of the sampling unit.

Pattern detection in populations

For closed study regions in which point-plant distances are measured, the number of points used should be equal to or greater than the number of plants (Diggle & Matérn 1980; Ripley 1981). If

plant-to-plant distances are measured, the interpretation of spatial pattern becomes more confident as the area sampled is increased (Galiano 1982). Increasing the area covered may result in greater environmental heterogeneity, however, rendering the interpretation of spatial pattern more difficult (Pielou 1962; Kenkel 1988a). The same argument holds for the quadrat-variance methods (Greig-Smith 1983): the larger the nested quadrat system, the more likely that patterns attributable to intraspecific interactions will be confounded by extraneous factors.

Pattern detection in communities

If the objective is trend-seeking, the empirical sample size should be such that an ordination or classification does not change appreciably with any further addition of sampling units. The empirical sample size at which the results stabilize should therefore be determined. In ordination, stability implies that the scattergram trends do not change when more sampling units are added. In classification, stability implies that the group structure remains unaltered with the addition of sampling units. Analytical techniques such as jackknifing and bootstrapping are useful in assessing the stability of results. The effect of including additional species in a reference ordination was examined by Goff (1975), but corresponding studies of plot ordinations have not been undertaken. Podani (1986) examined the extent to which a classification of plots changed with increasing empirical sample size. Changes were found to be more substantial at the six-cluster than the three-cluster level, suggesting instability (or non-classifiability) in the first case and reasonable stability in the second. The analytical method (clustering algorithm, resemblance measure) used may also be important in determining the empirical sample size beyond which a classification of species stabilizes (Podani 1987).

When typification is the major objective, the problem of determining empirical sample size is potentially more complex. Typification often attempts to accomplish two goals simultaneously:

the delineation of vegetation types, and the description of each of these types. A sample size which results in any of the delineated types being represented by only one or a few sampling units is therefore inadequate; the empirical sample size must be large enough to ensure both a stable classification, and adequate representation of all delineated vegetation types. Ideally, sampling under a typification objective should involve two stages: the first to define and locate vegetation types, and the second to intensively sample within each type for estimation purposes.

Selection or arrangement of sampling units

Univariate estimation

In the discrete sampling universe, unbiased estimates can only be obtained through random selection. In a complete random design, all units have an equal chance of being selected. Precision can often be increased through various 'restricted' randomization procedures. In stratified random (frame) sampling, the strata or sub-populations are defined *a priori* by the user. Greater precision is expected when strata are homogeneous within and heterogeneous among. Note that in order to obtain an estimate of the sampling variance, at least two sampling units must be taken from each stratum. In systematic sampling, a pivotal individual is chosen at random and every r^{th} individual in the sampling frame is enumerated. This method, while convenient, does not give an unbiased estimate of the sampling variance. One way to overcome this problem is to perform stratified systematic sampling, in which two independent systematic samples are taken from each stratum (Cochran 1977, p. 227).

For the continuous sampling universe, randomization is again required. In this case, however, it is the arrangement of the sampling units within the study area that is randomized. 'Plot-less' sampling (in which points are located and distances to the closest plant measured) requires full randomization to estimate density (Greig-Smith 1983, p. 48). With plots, complete random-

ization is often difficult to achieve as random points are not easily located in the field (Laferrière 1987). Under these circumstances systematic sampling may be more easily implemented: again, the method should be modified so as to permit unbiased estimation of the sampling variance. Note that if the sampling interval coincides with periodic variation in vegetation (spatial autocorrelation), the sample will not be representative of the sampling universe (Finney 1950; Greig-Smith 1983, p. 23; De Vries 1986, p. 25). Restricted randomization designs involving both systematic and random components may therefore be preferred. Studies comparing the efficiency of various arrangements for estimation purposes include Hasel (1938), Pechanec & Stewart (1940), Finney (1948, 1950), Goodall (1952), Bourdeau (1953), Kulow (1966), Yandle & Wyant (1981), and Diggle & ter Braak (1982). Smartt & Grainger (1974) address the problem of determining the optimal arrangement of units when community types are already defined and interest lies in estimating the relative proportion of types in an area. Smartt (1978) outlines a flexible model for optimal allocation of sampling units.

Multivariate estimation

Individual sampling units must be selected at random in order to obtain unbiased estimates of variable means and their intercorrelations. We are unaware of any studies dealing with multivariate estimation under restricted random (stratified or systematic) sampling designs.

Pattern recognition in populations

The nature of the sampling universe determines two different approaches to the analysis of pattern in a single species. For the discrete universe, individuals may be selected at random and plant-to-nearest neighbor distances measured. Clark & Evans (1954) suggested this method as a test of departure from spatial randomness. Alternatively the plant closest to a random point may be

chosen, and the distance between that plant and its nearest neighbor measured (e.g. Cox & Lewis 1976). Diggle (1979a) suggested examining the complete distribution function of nearest-neighbor distances. For mapped point patterns, the combined count-distance method of Ripley (1977) offers the most penetrating analysis (Prentice & Werger 1985; Kenkel 1988a). These and other recently developed techniques are reviewed by Upton & Fingleton (1985, pp. 74–90).

For a continuous sampling universe, point-to-plant methods represent the simplest situation. Methods involving randomly located points have been suggested by Pielou (1959) and Mountford (1961). Others have suggested that a regular grid of points (Diggle 1979b) or a stratified arrangement (Ripley 1981) may be preferred.

Analysis of spatial pattern using plots has yielded a wide variety of designs. In the simplest case, plots are randomly located within the study area and departures from randomness assessed using various aggregation indices (Goodall & West 1979; Carpenter & Chaney 1983). While non-contiguous quadrats are sometimes used in ecological studies (Upton & Fingleton 1985, p. 65), contiguous units in a regular grid or long transect are more commonly employed (e.g. Greig-Smith 1952; Hill 1973; Ludwig & Goodall 1978; Galiano 1985). A combination of random and systematic designs was used by Goodall (1974). Renshaw & Ford (1984) point out that a grid of plots is preferable to a long transect, since only the former can detect directional (anisotropic) trends in spatial pattern.

Pattern detection in community studies

The distinction between the discrete and continuous sampling universe is important throughout the following discussion. If discrete, the sampling units are selected from the sampling frame. For the continuous case a map of the study area is normally required to locate sampling units in the field. Edge effect problems may arise in the latter case.

Sampling units may be randomly located within

the study area. However, random points may in practice be difficult to locate in the field. Furthermore, random sampling generally results in a relatively inequitable distribution of sampling units over the study area, which is disadvantageous in survey and mapping studies. To overcome these problems, the study area may first be stratified, and independent random samples taken from each stratum. Strata can be defined in various ways: the simplest involves gridding the study area into equal-sized units, and taking a random sample from within each. Alternatively, an underlying factor (e.g. elevation categories) can be used to define strata. If typification is the objective, however, circularity of argument must be avoided: thus, strata should not be defined using dominant vegetation if the objective is to define and describe vegetation types (Orlóci 1978).

Systematic sampling is useful in vegetation mapping, since the sampling units are equally spaced. For this reason it is also the method of choice in direct gradient analysis studies. Systematically placed sampling units are also more readily located in the field.

In community pattern analysis, an ordination procedure is often employed to obtain a univariate seriation of plots. The ordination scores are then used as input to one of the univariate spatial pattern detection methods outlined above. Ecological examples include Bouxin (1983), Galiano (1983), Castro *et al.* (1986), Gibson & Greig-Smith (1986), and Kenkel (1988b).

Plot size

Univariate estimation

Plot size under an estimation objective will depend upon the spatial pattern of the sampling universe. If random, any plot size will give an equally precise estimate of the mean (or total). Many plant species show a clumped pattern, however, and as a consequence plot size determination will be important in minimizing the sampling variance. Studies reporting on the effect

of plot size on precision include Freese (1961), Van Dyne *et al.* (1963), Wiegert (1967), and Ram Babu *et al.* (1981).

It is useful to distinguish two cases. In the first, the total amount of material sampled is fixed by practical constraints (typically expressed as a cost function). Examples include biomass studies, and studies employing benthic grabs or soil cores. In all these cases, there is a fixed upper limit on the amount of material (A) which can be processed. Under such circumstances, the smallest possible area or volume per sampling unit will normally result in the most precise estimate of the mean (Green 1979, p. 132; Elliott 1983, p. 128). The reason for this is simple: we note that the fixed total $A = a \times n$, where n is the sample size and a is the size (area or volume) of each sampling unit. Clearly, the smaller the sampling unit, the larger the sample size. Because the standard error of the mean $S_{\bar{x}} = [S^2/n]^{1/2}$, precision can always be improved by making n larger (that is, making a smaller). In practice, the size of the organism sampled, edge effects, and other factors will determine the smallest sampling unit size possible (Bigwood & Inouye 1988).

There are other circumstances, however, in which the sample size n is fixed, but the size of the plot can vary within certain limits. For example, consider the determination of percent cover in each of n randomly located plots within a large study area. The location of plots in the field is a relatively time-consuming exercise. Once located, however, the time taken to estimate species percent cover is largely independent of plot size, at least within certain size limits. Under these circumstances, a plot size should be chosen which minimizes the sampling variance (the between-plot variation). Such a plot will normally be somewhat larger than the mean clump size. A further advantage of using a larger plot size is that a greater proportion of the study area is sampled.

For diversity and evenness estimation, increasing plot size has the same effect as increasing the sample size, since it is the total area sampled which determines the precision of estimates (Kwiatkowska & Symonides 1986).

Multivariate estimation

Here the objective is to select a plot size that simultaneously minimizes the variances of all variables. In addition, the pairwise associations between species should be minimized (c.f. Goodall 1961). This is most easily accomplished by minimizing the covariance structure of the cross-products matrix between variables. A plot size which simultaneously minimizes the variances of species, and their covariances, will therefore be optimal for estimation of mean or total values in community studies.

Pattern recognition in populations

The scale of spatial pattern of a species can be detected by locating plots of different sizes within a study area. One possibility would be to take independent samples using different plot sizes, but this is generally not practical in the field situation. An alternative is to begin with a regular grid of sufficiently small units, and then to successively fuse adjacent units to obtain larger plots or blocks (Greig-Smith 1952). Similar nested systems have been suggested for examining transect data (Hill 1973; Goodall & West 1979). Such data can then be subjected to statistical analysis to detect departures from randomness (Upton & Fingleton 1985, pp. 46–48). Under some circumstances departures from randomness in non-random populations are largely independent of plot size. Pielou (1977, p. 136) indicates that this may occur when compact clumps are widely spaced, or when mosaics of differing density are present.

Pattern recognition in community studies

Ideally the analysis of community pattern in the continuous universe should involve the use of a series of plot sizes, since results are dependent on the size of the plot used. For example, interspecific species associations are known to vary greatly as plot size is changed (Greig-Smith 1983), and as

a result species classifications and ordinations are also prone to change (Fekete & Szöcs 1974; Matthews 1978). The dependence of species covariance structure on plot size may indicate a characteristic area of the community in terms of maximum covariance ('multiple pattern analysis', Noy-Meir & Anderson 1971). Similarly, a range of plot sizes is required to detect an area at which maximum differences occur between random pairs of plots (expected resemblance, Podani 1984a, b). Concepts of florula diversity and related information theory functions (Juhász-Nagy & Podani 1983) are also inevitably linked to a screening process involving increases in plot size. This holds true for other structural variables as well.

Results are also scale-dependent when the objective is the classification and ordination of plots (Kershaw 1961; Noy-Meir *et al.* 1970). This problem is particularly complicated in typification studies. The recognition of vegetation types is clearest when plots not belonging to the same type are maximally different. At the same time, plots belonging to the same type should be as similar as possible. Clearly these are potentially conflicting objectives. Ideally, sampling should proceed in two stages. In the first, a plot size should be chosen which maximizes between-plot variation in order to best discriminate between types. In the second stage, a different plot size should be used within each vegetation type to obtain the most precise community estimates.

It would perhaps be instructive to mention here the concept of minimal area, as much confusion exists in the literature concerning its utility and purpose. The minimal area method attempts to find the smallest plot size necessary to characterize the species richness of a community (species-area curves, Mueller-Dombois & Ellenberg 1974): that is, it determines the spatial scale of species richness (Dietvorst *et al.* 1982). The method is useful for plot size determination in phytosociology, where the objective is to describe a previously-defined community as a single 'picture' or relevé. It is inappropriate as a method for determining plot size in probabilistic sampling, however.

Plot shape

Univariate estimation

Like plot size, the shape of a plot is dependent on the spatial pattern of a species. Precision of estimates is unaffected by plot shape if the spatial pattern is random, though elongated plots will increase edge effects. Elongated plots are preferred when the pattern is aggregated, however, since the plot will tend to include both high and low density regions and thereby reduce the sampling variance. This is well documented in studies estimating yield (e.g. Kalamakar 1932; Justesen 1932; Van Dyne *et al.* 1963), abundance (Clapham 1932), and various forest parameters (Hasel 1938; Bormann 1953; Freese 1961).

Elongated plots are also preferred when estimating diversity and related textural variables, since more diverse regions of the study area are likely to be included. Hahn (1982) has explored some possibilities for edge effect correction in estimating diversity.

Multivariate estimation

Optimization of plot shape is a complex problem in this case. Whereas the precision of estimates for a given species will likely increase when the plot is elongated, the species covariances are subject to unpredictable changes. Examination of several plot shapes in a pilot study to select a shape offering the best compromise between precision and covariance minimization is therefore recommended. We are unaware of any studies that have examined this problem.

Pattern detection in populations

Clearly, pattern detection is most effective when plots are of the same size and shape as the mean individual clump size. This could be found by varying plot size and shape in a factorial design, and finding the size-shape combination at which departures from randomness are maximized.

Care must be taken to ensure that elongated plots include only a single clump, however, since otherwise different hierarchical levels of spatial pattern will be confounded.

Pattern detection in community studies

Whereas inclusion of heterogeneous areas in the same plot is advantageous in parameter estimation, elongated sampling units may have detrimental effects in community pattern studies. This is because interspecific associations detected using elongated plots are artifactual and potentially misleading (Pielou 1977; Greig-Smith 1983). The use of elongated plots should also be avoided when measuring expected resemblance, florula diversity and related structural variables (Podani 1984b; Bartha & Horváth 1987). Since square or circular plots are least likely to overlap community boundaries, they are also recommended for ordination and classification studies. A further advantage of isodiametric plots in community pattern detection is that edge effect problems are minimized. In practice, square plots may be easier to set up and enumerate than circular ones.

Concluding remarks

We have emphasized that sampling decisions must be consistent with the objectives of an investigation. In addition, we suggest that many sampling problems have received a cursory treatment in the ecological literature. Familiarity with the dichotomous sampling alternatives outlined here will hopefully help ecologists in making important sampling decisions.

This review indicates that there remain a number of unresolved questions regarding sampling in ecology. Our hope is that some of the issues we have raised will be explored further, and that ultimately specific 'optimal' sampling strategies will be determined for various objectives. It would also be desirable to determine the relative importance of sampling in ecological studies: that is,

how important are sampling decisions relative to decisions regarding choices of data type and standardization, the resemblance measure used, and the scaling or clustering procedure employed? The effect of sampling on the stability of results is also deserving of greater attention.

Acknowledgements

We thank D. W. Goodall, R. H. Green, and L. Orlóci for their stimulating discussions on various sampling topics. Two anonymous reviewers offered useful comments. Support in the form of a Soros Foundation (New York-Budapest) travel grant to J. Podani to finance a visit to Winnipeg, Canada is gratefully acknowledged. This research was supported by Natural Sciences and Engineering Research Council of Canada grant A-3140 to N.C. Kenkel.

References

- Barkman, J. J. 1979. The investigation of vegetation texture and structure. In: Werger, M. J. A. (ed.), *The study of vegetation*. pp. 125–160. Junk, The Hague.
- Bartha, S. & Horváth, 1987. Application of long transects and information theoretical functions to pattern detection. I. Transects versus isodiametric sampling units *Abstr. Bot.* 11: 9–26.
- Bigwood, D. W. & Inouye, D. W. 1988. Spatial pattern analysis of seed banks: an improved method and optimized sampling. *Ecology* 69: 497–507.
- Bormann, G. E. 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.
- Bouxin, G. 1983. Multi-scaled pattern analysis: an example with savanna vegetation and a proposal for a sampling design. *Vegetatio* 52: 161–169.
- Braun-Blanquet, J. 1964. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. 3. Aufl. Springer, Wien.
- Carpenter, S. R. & Chaney, J. E. 1983. Scale of spatial pattern: four methods compared. *Vegetatio* 53: 153–160.
- Castro, I., Sterling, A. & Galiano, E. F. 1986. Multi-species pattern analysis of Mediterranean pastures in three stages of ecological succession. *Vegetatio* 68: 37–42.
- 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. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35: 445–453.
- Cochran, W. G. 1977. *Sampling techniques*. 3rd ed. J. Wiley & Sons, New York.
- Cox, T. F. & Lewis, T. 1976. A conditioned distance ratio method for analyzing spatial patterns. *Biometrika* 63: 483–491.
- Croy, C. D. & Dix, R. L. 1984. Notes on sample size requirements in morphological plant ecology. *Ecology* 65: 662–666.
- De Vries, P. G. 1986. *Sampling theory for forest inventory*. Springer, Berlin.
- 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.
- Digby, P. G. N. & Kempton, R. A. 1987. *Multivariate analysis of ecological communities*. Chapman & Hall, London.
- Diggle, P. J. 1979a. Statistical methods for spatial point patterns in ecology. In: Cormack, R. M. & Ord, J. K. (eds), *Spatial and temporal analysis in ecology*. pp. 95–150. Inter. Coop. Publ. House, Burtonsville, USA.
- Diggle, P. J. 1979b. On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* 35: 87–101.
- Diggle, P. J. & Matérn, B. 1980. On sampling designs for the study of point-event nearest neighbour distributions in R^2 . *Scand. J. Statist.* 7: 80–84.
- Diggle, P. J. & ter Braak, C. J. F. 1982. Point sampling of binary mosaics in ecology. In: Ranney, B. (ed.), *Statistics in theory and practice. Essays in the honour of Bertil Matérn*. pp. 107–122. Swedish Univ. Agric. Sci., Section of Forest Biometry, S-901 83, Umeå, Sweden.
- Elliott, J. M. 1983. Some methods for the statistical analysis of samples of benthic invertebrates. *Freshwater Biol. Assoc. Sci. Publ.* 25.
- Fekete, G. & Szöcs, Z. 1974. Studies on interspecific association processes in space. *Acta Bot. Acad. Sci. Hungary* 20: 227–241.
- 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.
- Freese, F. 1961. Relation of plot size to variability: an approximation. *J. Forestry* 58: 679.
- Galiano, E. F. 1982. Pattern detection in plant populations through the analysis of plant-to-all-plants distances. *Vegetatio* 49: 39–43.
- Galiano, E. F. 1983. Detection of multi-species patterns in plant populations. *Vegetatio* 53: 129–138.
- Galiano, E. F. 1985. The small-scale pattern of *Cynodon dactylon* in Mediterranean pastures. *Vegetatio* 63: 121–127.
- Gauch, H. G. 1982. *Multivariate analysis of community data*. Cambridge Univ. Press, Cambridge.
- Gibson, D. J. & Greig-Smith, P. 1986. Community pattern analysis: a method for quantifying community mosaic structure. *Vegetatio* 66: 41–47.
- Goff, F. G. 1975. Comparison of species ordinations resulting from alternative indices of interspecific association and different number of included species. *Vegetatio* 31: 1–14.

- Goodall, D. W. 1952. Some considerations in the use of point quadrats for the analysis of vegetation. *Aust. J. Sci. Res., Ser. B* 5: 1–41.
- 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. 1974. A new method for the analysis of spatial pattern by random pairing of quadrats. *Vegetatio* 29: 135–146.
- Goodall, D. W. & West, N. E. 1979. A comparison of techniques for assessing dispersion patterns. *Vegetatio* 40: 15–27.
- Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. J. Wiley & Sons, New York.
- Greig-Smith, P. 1952. The use of random versus contiguous quadrats in the study of the structure of plant communities. *Ann. Bot. Lond. N.S.* 16: 293–316.
- Greig-Smith, P. 1983. Quantitative plant ecology. 3rd ed. U. California Press, Berkeley.
- Hahn, I. 1982. Einige Probleme der Probeentnahme bei der Schätzung der Arten und Individuendiversität. II. Eine mögliche Individualzahnkorrektur. *Bot. Közlem., Budapest* 69: 59–70 (Hungarian with German summary).
- Hasel, A. A. 1938. Sampling errors in timber surveys. *J. Agric. Res.* 57: 713–736.
- Hill, M. O. 1973. The intensity of spatial pattern in plant communities. *J. Ecol.* 61: 225–235.
- 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.
- Kalamakar, R. J. 1932. Experimental error and the field plot technique with potatoes. *J. Agric. Sci.* 22: 373–385.
- Kenkel, N. C. 1988a. Pattern of self-thinning in jack pine: testing the random mortality hypothesis. *Ecology* 69: 1017–1024.
- Kenkel, N. C. 1988b. Spectral analysis of hummock-hollow pattern in a weakly minerotrophic mire. *Vegetatio* 78: 45–52.
- Kershaw, K. A. 1961. Association and covariance analysis of plant communities. *J. Ecol.* 49: 643–654.
- Knapp, R. (ed.) 1984. Sampling methods and taxon analysis in vegetation science. Junk, The Hague.
- Kulow, D. L. 1966. Comparison of forest sampling designs. *J. Forestry* 64: 469–474.
- Kwiatkowska, A. J. & Symonides, E. 1986. Spatial distribution of species diversity indices and their correlation with plot size. *Vegetatio* 68: 99–102.
- Laferrrière, J. E. 1987. A central location method for selecting random plots for vegetation surveys. *Vegetatio* 71: 75–77.
- 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.
- Ludwig, J. A. & Reynolds, J. F. 1988. Statistical ecology. J. Wiley & Sons, New York.
- Matthews, J. A. 1978. An application of non-metric multidimensional scaling to the construction of an improved species plexus. *J. Ecol.* 66: 157–173.
- Mountford, M. D. 1961. On E. C. Pielou's index of non-randomness. *J. Ecol.* 49: 271–275.
- Mueller-Dombois, D. & Ellenberg, H. 1974. Aims and methods of vegetation ecology. J. Wiley & Sons, New York.
- Noy-Meir, I. & Anderson, D. J. 1971. Multiple pattern analysis, or multiscale ordination: towards a vegetation hologram? In: Patil, G. P., Pielou, E. C. & Waters, W. E. (eds), Many species populations, ecosystems, and systems analysis. pp. 207–225. Penn. State U. Press, University Park, Pa.
- 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.
- Orlóci, L. 1978. Multivariate analysis in vegetation research. 2nd ed. Junk, The Hague.
- Pechanec, J. F. & Stewart, G. 1940. Sagebrush-grass range sampling studies: size and structure of sampling units. *J. Amer. Soc. Agron.* 32: 669–682.
- Pielou, E. C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. *J. Ecol.* 47: 607–613.
- Pielou, E. C. 1962. The use of plant-to-neighbour distances for the detection of competition. *J. Ecol.* 50: 357–367.
- Pielou, E. C. 1977. Mathematical ecology. J. Wiley & Sons, New York.
- Podani, J. 1984a. Spatial processes in the analysis of vegetation: theory and review. *Acta Bot. Hung.* 30: 75–118.
- Podani, J. 1984b. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. *Acta Bot. Hung.* 30: 403–425.
- Podani, J. 1986. Comparison of partitions in vegetation studies. *Abst. Bot.* 10: 235–290.
- Podani, J. 1987. Computerized sampling in vegetation studies. *Coenoses* 2: 9–18.
- Prentice, I. C. & Werger, M. J. A. 1985. Clump spacing in a desert dwarf shrub community. *Vegetatio* 63: 133–139.
- Ram Babu, M., Agarwal, 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.
- Renshaw, E. & Ford, E. D. 1984. The description of spatial pattern using two-dimensional spectral analysis. *Vegetatio* 56: 75–85.
- Ripley, B. D. 1977. Modelling spatial pattern. *J. Royal Stat. Soc., Series B* 39: 172–212.
- Ripley, B. D. 1981. Spatial statistics. J. Wiley & Sons, New York.
- Scagel, R., el-Kassaby, Y. A. & Emanuel, J. 1985. Assessing sample size and variable number in multivariate data, with specific reference to cone morphology variation in a population of *Picea sitchensis*. *Can. J. Bot.* 63: 232–241.
- Smartt, P. F. M. 1978. Sampling for vegetation survey: a

- flexible systematic model for sample location. *J. Biogeogr.* 5: 43–56.
- Smartt, P. F. M. & Grainger, J. E. A. 1974. Sampling for vegetation survey: some aspects of the behaviour of unrestricted and stratified techniques. *J. Biogeogr.* 1: 193–206.
- Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. & Asok, C. 1984. *Sampling theory of surveys with applications*. Iowa State Univ. Press, Ames.
- Upton, G. J. G. & Fingleton, B. 1985. *Spatial data analysis by example*. Vol. 1. Point pattern and quantitative data. J. Wiley & Sons, New York.
- 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.
- Wiegert, R. G. 1967. The selection of an optimum quadrat size for sampling the standing crop of grasses and forbs. *Ecology* 43: 125–129.
- Wilks, S. S. 1962. *Mathematical statistics*. J. Wiley & Sons, New York.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia* 50: 296–302.
- Yandle, D. O. & Wyant, H. V. 1981. Comparison of fixed-radius circular plot sampling with simple random sampling. *Forest Sci.* 27: 245–252.