

Plot size and estimation efficiency in plant community studies

Kenkel, N. C.^{1,2*} & Podani, J.^{1,3}

¹International Centre for Environmental and Marine Sciences and Technologies, I.C.T.P., Via Beirut 7, I-34100, Trieste, Italy; ²Permanent address: Department of Botany, University of Manitoba, Winnipeg, Manitoba R3T 2N2 Canada; ³Permanent address: Department of Plant Taxonomy and Ecology, L. Eötvös University, Ludovika tér 2, Budapest, H-1083, Hungary;

*Tel. +1 204 474 9368; Fax +1 204 261 8474; E-mail MCILRTH@CCM.UMANITOBA.CA

Abstract. We examine the effects of changing plot size on parameter estimation efficiency in multivariate (community-level) ecological studies, where estimation efficiency is defined in terms relating to the statistical precision of estimates of all variables (e.g. species) in a data set. Three 'efficiency criteria' for multivariate estimation are developed, and the relationship between estimation efficiency and plot size examined using three field data sets (deciduous understory, coniferous understory, and mire vegetation) from central Canada.

For all three communities, estimation efficiency was found to increase monotonically with increasing plot size. However, relative gains in efficiency at larger plot sizes were offset by substantial increases in sampling effort (enumeration time per plot). Our results indicate that the largest plot size possible, given the constraints of time, should be used for parameter estimation in plant communities. Also, plots that are larger than the mean patch size should be utilized when sampling heterogeneous vegetation.

Keywords: Covariance; Eigenanalysis; Pattern; Sampling; Spatial process; Understory; Variance.

Nomenclature: Scoggan (1978).

Abbreviations: PCA = Principal Components Analysis; VAR = variance.

Introduction

It is important to distinguish between the sampling objectives of pattern recognition and parameter estimation when using plots to sample plant communities (Kenkel, Juhász-Nagy & Podani 1989). Pattern recognition refers to the detection of the intensity and scale of variation in populations and communities, and includes the delineation of vegetation types and descriptions of

vegetation variation (Orlóci 1978). This paper is concerned with parameter estimation, which occurs when a given population or community is 'characterized' in some way. Examples from vegetation science include the determination of species richness, and mean species abundance or density, within a plant community.

Vegetation scientists have traditionally used the minimal area approach when characterizing plant communities. Minimal area is generally defined as the plot size beyond which species richness fails to increase, or increases only slightly. It is often determined through interpretation of a species - area curve, but this approach is fraught with problems (Barkman 1989). Gounot & Calléja (1962), and later Moravec (1973), suggested instead using similarity analysis (based on presence-absence data) to determine minimal area. Here, a plot of given size is said to be 'representative' if its floristic composition is similar to that of other plots (of the same size) taken from the same community (Barkman 1989). Quantitative minimal area is defined when floristic similarity is determined using quantitative rather than presence-absence data (Roux & Rieux 1981; Dietvorst, van der Maarel & van der Putten 1982). The implicit objective of all minimal area approaches lies in determining the size of a plot necessary to obtain a representative 'picture' of a plant community, rather than in measuring the variation of abundance estimates. In this paper we focus on sampling procedures for statistical estimation, in which interest lies both in parameter estimation and determination of the parameter's variability.

Several investigators have examined the relationship between the statistical precision of sample estimates and plot size. However, these studies have only considered the population or univariate case (i.e. the estimation of a single variable, such as mean density of a species); corresponding investigations for the community or multivariate case (i.e. when a single sample is

used to obtain estimates of more than one variable) are lacking. For both univariate and multivariate studies, it is important to consider any implicit sampling constraints on both plot and sample size (in this paper, sample size refers to the total number of plots enumerated, and plot size to the areal extent of each plot). In some studies, the total area or volume that can be sampled is fixed: examples include biomass harvest studies, and those utilizing soil cores (van Dyne, Vogel & Fisser 1963; Zeide 1980). In these cases smaller plots (or soil core volumes) generally result in increased statistical precision, since plot and sample size are inversely related (Green 1979, p. 132; Benoit, Kenkel & Cavers 1989). When interplot travel time is great, however, this rule does not hold; Zeide (1980) describes a method for optimal plot size determination (univariate case only) in such situations.

In vegetation science, it is more common for plot size to be largely independent of sample size, and for time constraints to determine the sample size. For the population (univariate) case, the optimal plot size for estimation is dependent upon the spatial distribution of the species being enumerated (Bormann 1953; Wiegert 1962). For a randomly distributed species, all plot sizes will result in equally precise estimates. For a species showing a clumped distribution, however, greater precision will result when plots somewhat larger than the mean clump size are used (Kenkel, Juhász-Nagy & Podani 1989). In multivariate (community-level) studies, it is possible for each species to have a different spatial distribution, and thus a different 'optimal' plot size. A sampling program using a different plot size for each species is clearly not feasible, however. What is required instead is a means for determining a single common plot size which results in (overall) high precision of estimates for all species.

In this paper we develop measures of estimation efficiency for multivariate studies. Greater estimation efficiency is implied whenever there is an overall increase in precision of estimates of all variables (species). We also explore the relationship between plot size and estimation efficiency using data from conifer understory, deciduous understory and boreal mire communities.

Measuring the efficiency of multivariate estimates

In developing measures of estimation efficiency for our multivariate studies, we make the following assumptions: (a) the community for which estimates are required has been previously delineated; (b) sampling units are plots located randomly within the community; and (c) the number of plots enumerated is fixed by time

constraints, but the researcher has some flexibility in the selection of plot size. With our objective and these assumptions in mind, we develop three estimation efficiency criteria.

Criterion 1

In univariate estimation, a lower sample variance (S^2) implies a more precise estimate of the mean or total (Cochran 1977; Kenkel, Juhász-Nagy & Podani 1989). In the multivariate extension, we consider the sum of variances as our first criterion:

$$C_1 = \sum_{i=1}^p S_i^2 \quad (1)$$

The summation is over the p species recorded in the sample. A lower sum of variances (C_1) implies greater efficiency in multivariate estimation.

Criterion 2

In the univariate case, the variance determines the sample size required to obtain an estimate within specific confidence limits (Eckblad 1991). For the multivariate case, our second criterion involves determination of the variability of the species variances:

$$C_2 = \text{VAR}\left(S_i^2\right). \quad (2)$$

Lower variability in species variances (C_2) implies that a smaller range of sample sizes is required for estimation purposes (Wiegert 1962; Eckblad 1991).

Criterion 3

Criteria C_1 and C_2 consider only the species variances. In the multivariate case, we must also consider the covariances, which can be thought of as measures of interspecific 'association'. Recall that in defining quantitative minimal area, a plot size is sought which results in high compositional similarity among replicate plots; such a plot size minimizes the detection of community 'patchiness', which in turn implies an overall reduction in species variances and interspecific associations (Dietvorst, van der Maarel & van der Putten 1982; Barkman 1989). This was recognized over 60 years ago by Ramenski, who defined minimal area as "that surface where the variance of abundance of all species is below a certain level" (quoted from Barkman 1989, p. 93). Similarly, Goodall (1953) developed a classification method in which vegetation groups are defined such that within-group species associations are

minimized. Suppression of interspecific associations is also desirable for multivariate estimation, since a plot size which detects patches or 'pockets of associations' within a community leads to inflation of species variances (Greig-Smith 1952, 1983). The extent of interspecific association can be measured as the amount of 'structure' in the species variance-covariance matrix S . Covariance 'structure' can be determined through an eigenanalysis of S (Gnanadesikan 1977; Orłóci 1978). Recall that the sum of eigenvalues equals the sum of the species variances (diagonal elements of S):

$$\sum_{i=1}^p \lambda_i = \sum_{i=1}^p S_i^2 \quad (3)$$

where summations are over the p species. When there is no covariance structure (all off-diagonal elements of S equal zero), the eigenvalues equal the variances. When covariance structure is high (species associations high, pairwise covariances large), the variance of eigenvalues will increase since the first few eigenvalues will be much larger than the others. In the extreme case, variables are completely redundant (covariances are at their maximum); the first eigenvalue is then

$$\lambda_1 = \sum_{i=1}^p S_i^2 \quad (4)$$

while the remaining $p - 1$ eigenvalues equal zero. It follows that a criterion for measuring the covariance 'structure' of S is the variance of the p eigenvalues:

$$C_3 = \text{VAR}(\lambda_i). \quad (5)$$

Note that zero eigenvalues are included in this calculation. For a given matrix S , C_3 is smallest when all the eigenvalues are equal and covariances are zero, and largest when equation (4) holds. A lower variance of eigenvalues (C_3) implies greater efficiency in multivariate estimation.

The relationship between plot size and estimation efficiency

To examine the relationship between plot size and multivariate estimation efficiency, we sampled three plant communities in central Canada. At each site, percent species cover estimates were obtained for each of seven plot sizes (a spatial process approach as described by Podani 1984).

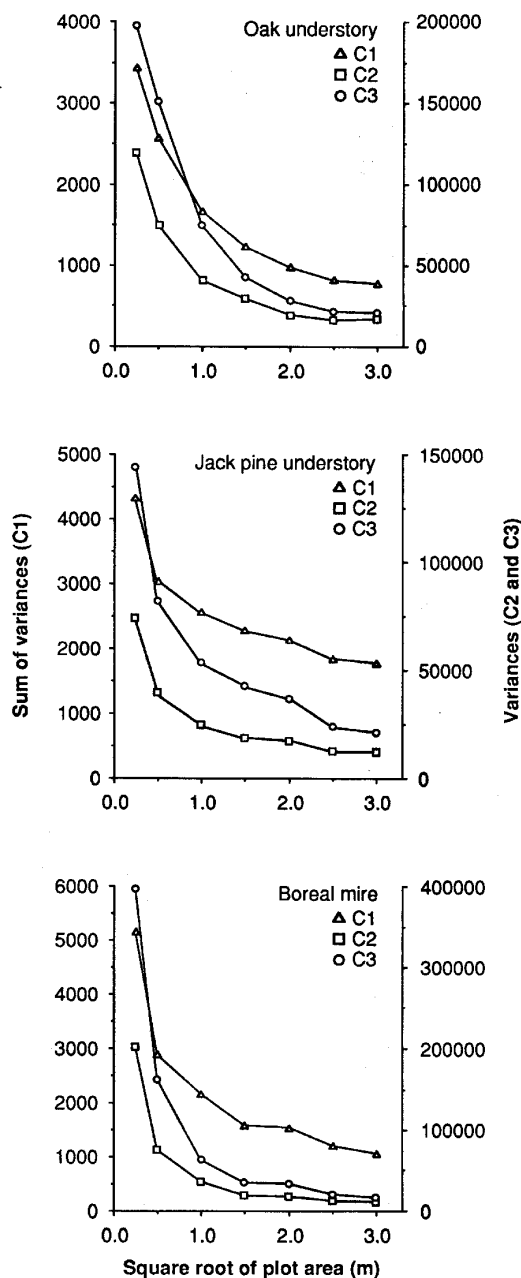


Fig. 1. The relationship between plot size and the criteria variables C_1 , C_2 and C_3 for three field data sets.

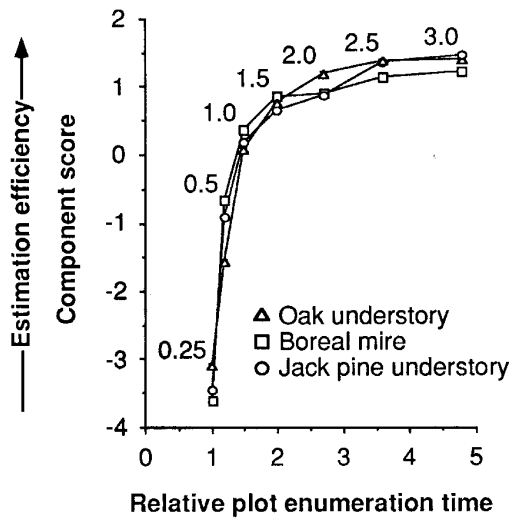


Fig. 2. The relationship between plot enumeration time (expressed relative to the time taken to enumerate the smallest plot) and estimation efficiency (scores on the first principal component axis). The values within the graph are plot sizes.

The plots were square with side lengths of 0.25, 0.50, 1.00, 1.50, 2.00, 2.50 and 3.00 m, and were nested at each of 20 random locations. The three communities are described below.

1. Oak forest understory vegetation: Delta, Manitoba (50° 11'N, 98° 23'W). This is a mature forest of *Quercus macrocarpa* and *Fraxinus pennsylvanica* abutting an oxbow of the Assiniboine River, 2 km south of Lake Manitoba. Soil is a slightly acidic clay-loam. There were 23 understory species, the dominants being *Aralia nudicaulis* and *Carex assiniboinensis*.

2. Jack pine understory vegetation: Elk Lake, Ontario (47° 50'N, 80° 27'W). This is a monodominant, fire-regenerated stand of *Pinus banksiana*. Soil is a medium-textured acidic sand. There were 18 understory species, the dominants being *Vaccinium angustifolium*, *Vaccinium myrtilloides*, and *Kalmia angustifolia*.

3. Mire vegetation: Star Lake, Manitoba (49° 45'N, 95° 14'W). This is a weakly minerotrophic acidic mire (organic soil) dominated by *Sphagnum* spp. and *Chamaedaphne calyculata*. There were 18 species. The area sampled had a somewhat hummocky microtopography and lacked trees, though *Picea mariana* occurred nearby.

Computations

For each data set, a species variance-covariance matrix S was computed for each plot size, and the sum and variance of the species variances calculated. Eigenanalysis was also performed on each matrix and the variance of eigenvalues determined.

Results

For all data sets, the three criteria ($C_1 - C_3$) show a monotonic decrease with increasing plot size (Fig. 1). Decreases are most pronounced at the smaller plot sizes, indicating a sensitivity of smaller plots to small-scale vegetation 'patchiness'. While no absolute minima were detected, the results do indicate that decreases in all three criteria become less pronounced with increasing plot size.

Incorporation of sampling effort

Determination of sampling effort

Field determinations of sampling effort were obtained for the three study sites by measuring the amount of time taken for two persons to set up and enumerate a single plot of specified size. Enumeration times for the three study sites proved to be very similar and were therefore averaged.

Estimation Efficiency Index

In order to estimate sampling efficiency we require a composite 'Estimation Efficiency Index' which utilizes all three of the criteria variables (C_1 , C_2 and C_3). Because the criteria variables are measured on different scales, a composite index cannot be obtained through simple averaging. Instead we computed pairwise correlations between the criteria variables over the seven plot sizes, and subjected the resulting correlation matrix to a Principal Components Analysis (PCA). This is justified since the three criteria are highly correlated across plot size (see Fig. 1); indeed, for all three data sets > 99.5% of the total variation was explained by the first principal component. The scores of the plot sizes on the first principal component defined the composite index of 'estimation efficiency'.

Results

For each data set, the relationship between estimation efficiency and sampling effort (plot enumeration time) is monotonic increasing (Fig. 2). Thus greater

estimation efficiency can always be achieved by increasing plot size. However, the results also indicate that for larger plot sizes any increase in estimation efficiency is offset by substantially greater sampling effort.

Discussion

In both population and community studies, an increase in sample size invariably results in improved estimation efficiency (Eckblad 1991). The 'effective' sample size may be increased either by enumerating more plots, or by increasing the size of each plot. Smaller plots are more likely to detect small-scale vegetation patches; this results in higher species variances and covariances and a concomitant decrease in estimation efficiency. It follows that plots somewhat larger than the mean patch size will result in increased estimation efficiency.

This parallels results showing that the minimal area of a plant community becomes greater with increasing scale and heterogeneity of vegetation patches (Barkman 1989), and with increasing species richness or decreasing dominance (Dietvorst, van der Maarel & van der Putten 1982). In practice these are also important considerations in selecting a plot size for statistical estimation, since plant communities are rarely homogeneous. Compositional heterogeneity in vegetation may reflect environmental heterogeneity, or be the result of biotic factors such as clonal growth and interspecific species interactions.

Our results indicate that estimation efficiency in community-level studies can always be improved by increasing plot size. However, it is also shown that while large increases in estimation efficiency occur at small plot sizes, only slight increases in estimation efficiency are observed at larger plot sizes. When sampling effort (enumeration time per plot) is incorporated, the decrease in improvement of estimation efficiency with increasing plot size is particularly apparent (see Fig. 2). In practice, we recommend that researchers utilize the largest plot size possible given the constraints of sampling time and effort. When sampling heterogeneous vegetation, we recommend that plots that are somewhat larger than the mean size of vegetation patches be utilized.

We have not yet considered the effect of changing plot shape on estimation efficiency. For the univariate case, it has been shown that elongated (rectangular) plots result in more statistically precise estimates (Bormann 1953); the same result can be expected for community-level studies. This will occur since rectangular plots are more likely to 'cross over' than detect

vegetation patches. This has the result that a greater proportion of the community's species are included within a single plot (Kenkel, Juhász-Nagy & Podani 1989). Rectangular plots may therefore be particularly useful when sampling heterogeneous vegetation given an objective of abundance estimation. Further studies should be undertaken to study the relationship between estimation efficiency and plot size - shape in multivariate sampling.

Acknowledgements. We thank D.W. Goodall, R.H. Green, M. B. Dale, T. Czárán, I. Scheuring, A. McIlraith and an anonymous reviewer for their comments and suggestions. The Institute for Earth, Environmental and Marine Sciences and Technologies, Trieste (ICEM) and the Centro di Ecologia Teorica ad Applicata, Gorizia (CETA) provided research facilities during our stay in Italy. This research was supported by Natural Sciences and Engineering Research Council of Canada operating grant A-3140 to N. C. Kenkel. Contribution No. 220 from the University of Manitoba Field Station (Delta Marsh).

References

- Barkman, J. J. 1989. A critical evaluation of minimal area concepts. *Vegetatio* 85: 89-104.
- Benoit, D. L., Kenkel, N. C. & Cavers, P. B. 1989. Factors influencing the precision of soil seed bank estimates. *Can. J. Bot.* 67: 2833-2840.
- Bormann, G. E. 1953. The statistical efficiency of sample plot size and shape in forest ecology. *Ecology* 34: 474-487.
- Cochran, W. G. 1977. *Sampling techniques*, 3rd ed. J. Wiley, New York.
- 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.
- Eckblad, J. W. 1991. How many samples should be taken? *Bioscience* 41: 346-348.
- Gnanadesikan, R. 1977. *Methods for statistical analysis of multivariate observations*. J. Wiley, New York.
- Goodall, D. W. 1953. Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. *Aust. J. Bot.* 1: 39-63.
- Gounot, M. & Calléja, M. 1962. Coefficient de communauté, homogénéité et aire minimale. *Bull. Serv. Carte Phytogeogr.* 7B: 181-210.
- Green, R. H. 1979. *Sampling design and statistical methods for environmental biologists*. J. Wiley, 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. University of California Press, Berkeley, CA.
- Kenkel, N. C., Juhász-Nagy, P. & Podani, J. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195-207.

- Moravec, J. 1973. The determination of the minimal area of phytocoenoses. *Folia Geobot. Phytotax.* 8: 23-47.
- Orlóci, L. 1978. *Multivariate analysis in vegetation research*, 2nd ed. Junk, The Hague.
- Podani, J. 1984. Spatial processes in the study of vegetation: theory and review. *Acta Bot. Hung.* 30: 75-118.
- Roux, C. & Rieux, R. 1981. L'aire minimale des peuplements lichéniques saxicoles-calicoles. *Vegetatio* 44: 65-76.
- Scoggan, H. J. 1978. *Flora of Canada*. National Museums of Canada, Ottawa, Ont.
- 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. 1962. The selection of an optimum quadrat size for sampling the standing crop of grasses and forbs. *Ecology* 43: 125-129.
- Zeide, B. 1980. Plot size optimization. *For. Sci.* 26: 251-257.

Received 13 December 1990;
Revision received 26 June 1991;
Accepted 28 June 1991.