Biom. Praxim. (1989), 29, 173-184

ON SAMPLE SIZE OPTIMALITY IN ECOSYSTEM SURVEY

L. ORLÓCI and V. De PATTA PILLAR

Department of Plant Sciences
University of Western Ontario
LONDON, Canada N6A 5B7

SUMMARY

Sampling is discussed as a process. In this the sample structure evolves and attains increasing stability as the sample size n increases. The minimum n at which the sample structure begins to attain stability is suggested as a lower bound for optimal sample size in ecosystem survey. Concept and method are illustrated by examples: one concerned with determination of optimal sample size in a field survey, and the other with probing the sample for the adequacy of its size.

1. INTRODUCTION

It is interesting to note how far the standard statistical texts are willing to go with assumptions about the sampled medium. These texts present sampling theory in terms of an idealized medium whose units are discrete, unambiguous and well behaving in terms of occurrence, availability and the characteristics that they posses. The sampling objectives are kept simple, rarely more complex than an estimation of frequency or performance. These texts narrow the design to the point that sample size becomes the sole determinant of sampling optimality. In such a setup, sample size determination is a matter of balancing some variance related requirement (SAMPFORD, 1962) with the cost of sampling.

A characteristic of ecosystem survey, which sets it apart from a population census, is the ill-defined sampling environment. The medium is a complex aggregate of elements, such as the natural landscape. It exists spontaneously and has no natural units. The objectives specify complex tasks such as the discovery and description of structural patterns and pattern coincidences in vegetation and environment. It is not surprising that under these conditions even the elementary choices such as unit definition, unit

siting and unit number (sample size) become perplexing problems beyond possible solution based on the sampling tools of conventional statistics.

We are concerned with ecosystem sampling based on area units (quadrats, plots). In our sampling environment sample size[1] cannot be defined based on a variance criterion. We discuss the implications and present empirical optimality criterion.

2. PROCESS SAMPLING

The proposition in process sampling is reminiscent of POORE'S (1955, 1956) successive approximation approach and the flexible analysis of WILDI and ORLÓCI (1987). The term "process" conjures a view of sampling as a process in which step-by-step expansions of the sample are intricately tide to the evolution of sample structures and structural connections (JUHÁSZ-NAGY, P. and PODANI, 1983, PODANI, 1982, 1984, KENKEL, 1984, ORLÓCI , 1988, KENKEL, JUHÁSZ-NAGY, and PODANI, 1989). The evolving structures are monitored in concurrent data analysis based on which their stability is judged. When structural stability is detected the sampling stops.

3. SAMPLE SIZE IN QUADRAT-BASED ESTIMATION

Inferring the required sample size (n) from theoretical considerations related to the sampling variance SV, (S^2_x) is the variance of X, n the sample size and N the population size,

$$SV = \frac{S_x^2}{n}\left(1 - \frac{n}{N}\right) \qquad (1)$$

is a common statistical practice, but not recommended in ecosystem surveys. The reason is that SV will depend not only on sample size, but also upon influences unrelated to sample size, such as quadrat shape and quadrat size for which there is no provision in (1). Quadrat shape and size individually, or jointly, can decrease or increase SV through increased or decreased within-quadrat heterogeneity (GREIG-SMITH, 1983) at any given sample size. By this an anomalous situation is created in regard of which work by PODANI (1982, 1984) is highly relevant.

---

[1] Here, sample size is the number of quadrats in the sample.

4. SAMPLE SIZE IN QUADRAT-BASED DETECTION OF STRUCTURES

Manipulations of quadrat shape and size designed to reduce SV increase heterogeneity within quadrats. Yet, homogeneity is a necessary condition when structures and structural connections are sought. Clearly, estimation and structure seeking are contradictory objectives in quadrat-based sampling. Traditionally, phytosociologists use medium quadrat size (relative to the relevant scale of environmental variation), square quadrat shape, and preferential quadrat siting (Quadrats are laid in homogeneous environments, recognized by the homogeneity of vegetation. Often the determination of homogeneity is a matter of visual inspection), sometimes with a random (In one scenario, random choice decides the general location within environmental strata, but the quadrat is shifted as needed to an adjacent location to avoid excessive compositional heterogeneity among the four quadrats) element (e.g., ORLÓCI and STANEK, 1980), to improve within-quadrat homogeneity.

Sample size matters under any sampling objective, but its optimality depends on the objectives. Keeping in mind that in ecosystem surveys the objective is to recognize structural patterns and connections, it is sensible to use the stability of analytically mapped structures as our condition of sample size optimality. The minimum n at which the mapped structure begins to attain stability is the optimal sample size. We consider distance, entropy, information, and Eigenmappings.

4.1. DISTANCE MAPPINGS

An n x n symmetric matrix **D** of quadrat distances defines vegetation structure based on s species as variables. A second n x n symmetric matrix Δ of quadrat distances defines another structure based on t environmental variables. The similarity of the **D** and Δ configurations is a measure of the two structures' affinity.

The relationship of **D** and Δ evolves as sample size increases. We monitored this by a stress function

$$\sigma_{VE} = \sqrt{1 - \rho^2(\mathbf{D};\Delta)} \qquad (2)$$

in which $\rho(\mathbf{D};\Delta)$ is a product moment correlation. Other definitions are possible (e.g., SHEPARD and CARROLL, 1966).

4.2. DIVERSITY MAPPINGS

In a sample of s species and n quadrats, we define diversity structures for species and quadrats based on RÉNYI's entropy function,

$$H^{\alpha} = \frac{\ln \sum_{h=1}^{s} \sum_{j=1}^{n} p_{hj|T}^{\alpha}}{1 - \alpha} \quad \text{with} \quad p_{hj|T} = \frac{x_{hj}}{x_{..}} \tag{3}$$

(see ORLÓCI, 1978 , FEOLI, LAGONEGRO and ORLÓCI, 1984).

4.2.1. Structures involving species. The appropriate entropy partition is between species (bs) and within species (ws)

$$H^{\alpha} = H_{bs}^{\alpha} + H_{ws}^{\alpha} \tag{4}$$

in which

$$H_{bs}^{\alpha} = \frac{\ln \sum_{h=1}^{s} p_{h.}^{\alpha}}{1 - \alpha} \quad \text{with } p_{h.} = \frac{x_{h.}}{x_{..}} \tag{5}$$

$$H_{ws}^{\alpha} = \frac{\sum_{h=1}^{s} p_{h.}^{\alpha \cdot} \ln \sum_{j=1}^{n} p_{hj|s}^{\alpha}}{1 - \alpha} \tag{6}$$

$$\text{with } p_{h.}^{\alpha \cdot} = \frac{p_{h.}^{\alpha}}{\sum_{e=1}^{s} p_{e.}^{\alpha}} \text{ and } p_{hj|s} = \frac{x_{hj}}{x_{h.}}$$

4.2.2. Structures involving quadrats. The relevant entropy partition is between quadrats (bq) and within quadrats (wq)

$$H^{\alpha} = H_{bq}^{\alpha} + H_{wq}^{\alpha} \tag{7}$$

is the joint entropy (the same as in Eq. 3). Other definitions include:

$$H^{\alpha}_{bq} = \frac{\ln \sum_{j=1}^{n} p^{\alpha}_{.j}}{1 - \alpha} \tag{8}$$

$$H^{\alpha}_{wq} = \frac{\sum_{j=1}^{n} p^{\alpha'}_{.j} \ln \sum_{h=1}^{s} p^{\alpha}_{hj|Q}}{1 - \alpha} \tag{9}$$

$$\text{with } p_{.j} = \frac{X_{.j}}{X_{..}} ; \; p^{\alpha'}_{.j} = \frac{p^{\alpha}_{.j}}{\sum_{e=1}^{n} p^{\alpha}_{.e}} ; \; p_{hj|Q} = \frac{X_{hj}}{X_{.j}}$$

H^{α}_{ws} is the equivocation entropy of quadrats with respect to species and H^{α}_{wq} is the equivocation entropy of species with respect to quadrats (ORLÓCI and PILLAR, 1989).

4.3. INFORMATION MAPPINGS

The determining relation is RÉNYI's information function,

$$H^{\alpha}_{species;quadrats} = \frac{\ln \sum_{h=1}^{s} \sum_{j=1}^{n} \frac{p^{\alpha}_{hj|T}}{(p_{h.} p_{.j})^{\alpha-1}}}{\alpha - 1} \tag{10}$$

This is shared entropy, related to H^{α}, H^{α}_{bs} and H^{α}_{bq} (Eqs 3, 5, 8) in the manner of

$$H^{\alpha} = H^{\alpha}_{bs} + H^{\alpha}_{bq} - H^{\alpha}_{species;\,quadrats} \tag{11}$$

Since the number of species and the number of quadrats affect the entropy/information quantities, for comparative purposes they are expressed in relative terms (see ORLÓCI and PILLAR, 1989). An alternative type of relative quantity is Rajski's metric (0 and 1 limits):

$$d^{\alpha}_{species;\,quadrats} = 1 - \frac{H^{\alpha}_{species;\,quadrats}}{H^{\alpha}} \tag{12}$$

4.4. EIGENMAPPINGS

Given a square symmetric matrix of products **S** for s species based on the species quantities in n quadrats, the Eigenstructure is the set of Eigenvalues and Eigenvectors of **S**. As sample size increases, expectedly, an increasingly stabile Eigenstructure results.

5. OPTIMAL SAMPLE SIZE IN FIELD SAMPLING - EXAMPLE 1

In this example we demonstrate the use of structure stability to determine optimal sample size in field sampling. We use a data set from our Sub-Boreal recovery transect site at Elk Lake, Ontario, Canada. The vegetation is secondary 3 years after logging. The vegetation variables represent species cover/abundance. The environmental variables are elevation, exposure, slope, soil depth, and soil texture. The sampling is carried to 42 quadrats at which the number of species in the sample is 54. Each quadrat is 5 m. sq. Quadrat selection is random in steps of three quadrats between analyses. Figure 1 gives the stress graph for the distance structure, figure 2 for the diversity and information structures, and figure 3 for the Eigenstructure.

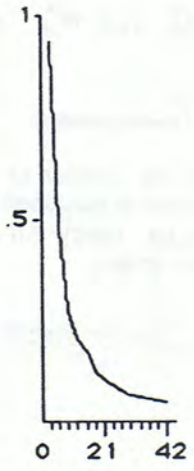

Figure 1: Changing relative stress (vertical scale, Eq. 2), comparing vegetation structure **D** and environmental structure Δ at increasing sample size (n, horizontal scale) in random sampling on the Elk Lake site. Although random sampling continues until 42 quadrats, structural stability is demonstrated already at about n = 18.

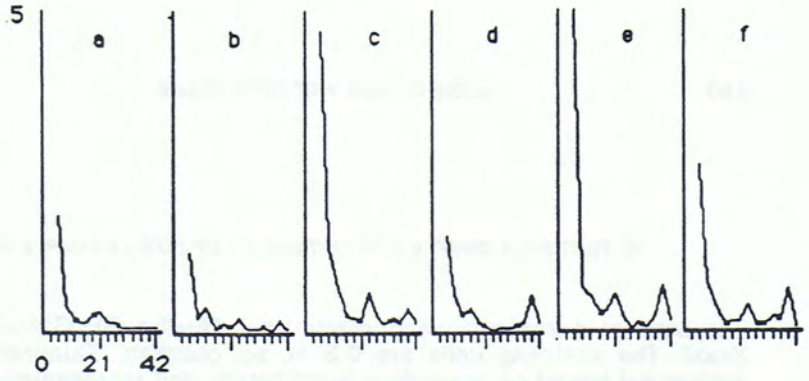

Figure 2: Changing relative stress (vertical scale), comparing diversity and information structures in the current sample of n quadrats and the preceding sample of n-3 quadrats in continued random sampling on the Elk Lake site. Horizontal scale indicates sample size (n). The structures begin to stabilize at about n = 12 in a, b and c and n = 18 in d, e and f. *Legend to diversity and information quantities:* a-joint (species and quadrats, Eq. 3); b - between species (Eq. 5); c - equivocation (quadrats conditional on species, Eq. 6); d - between quadrats (Eq. 8); e - equivocation (species conditional on quadrats, Eq. 9); f - Rajski's metric (species x quadrats, Eq. 12).

Figure 3: Changing relative stress (vertical scale), comparing the Eigenstructure in the current sample of n quadrats and the preceding sample of n-3 quadrats in continued random sampling on the Elk Lake site. Horizontal scale indicates sample size (n). The structure begins to stabilize at n = 15.

6. TESTING A SAMPLE FOR OPTIMALITY OF SIZE - EXAMPLE 2

The data set is from grassland vegetation in Guaíba, Rio Grande do Sul, Brazil. The sampling units are 0.5 m. sq. quadrats. Quadrat siting is preferential based on vegetation homogeneity and representativeness of subjectively observed patchiness. The sampling is carried to 60 quadrats at which the number of species is 165, reduced to 60 for analysis. The vegetation variables represent species cover/abundance. Environmental variables include relief position, grazing intensity, and 18 soil chemical and physical conditions. Detailed description of the site is given in PILLAR (1988) and PILLAR *et al.* (1989a,b). The sampling of the 60-quadrat sample is random in steps of three quadrats between analyses. The stress graphs are displayed in Figs. 4, 5, 6.

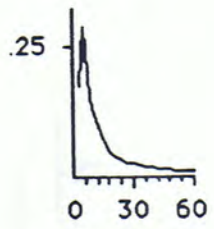

Figure 4: Changing relative stress (vertical scale, Eq. 2), comparing vegetation structure $\mathbb{D}$ and environmental structure Δ at increasing sample size (n, horizontal scale) in continued random sampling of the Guaíba sample. Structural stability is demonstrated already at about n = 15.

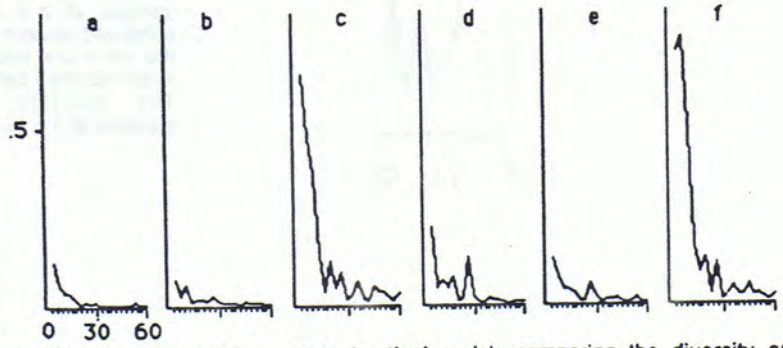

Figure 5: Changing relative stress (vertical scale), comparing the diversity and information structures in the current sample of n quadrats and the preceding sample of n-3 quadrats in continued random sampling of the Guaíba sample. Horizontal scale indicates sample size (n). The structures begin to stabilize at about n = 15 in a,b,c and e and at about n = 30 in d and f. See the legend to diversity and information quantities in the caption of Fig. 2.

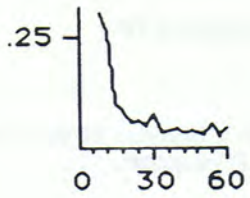

Figure 6: Changing relative stress (vertical scale), comparing the Eigenstructure in the current sample of n quadrats and the preceding sample of n-3 quadrats in continued random sampling of the Guaíba sample. Horizontal scale indicates sample size (n). The structure begins to stabilize at about n = 15.

7. CONCLUSIONS

We distinguished two major objectives in a survey. One was estimation and the other the discovery of biotic and environmental structures and structural connections. We assumed quadrat-based sampling and discussed the conditions of optimal sample size. We suggested that these conditions differ under the different objectives.

We investigated the effect of sample size on the stability of vegetation structures and structural connections under process sampling. In one case we were interested in field determination of sample size. In the other we tested the adequacy of the size of a given sample. We found that structures mapped by distance, entropy and information functions, and also the Eigenstructure, stabilize relatively early in the sampling process. The actual sample size was about two-folds larger than the minimum needed for structural stability in the Elk Lake sample, and three-folds larger in the Guaíba sample.

It is emphasized that structure is a specific sample property captured as an analytical mapping. The state of structural stability and the structures' sharpness are not related. Therefore, it is quite possible to have a stabile, but weakly defined structure.

RESUME

TAILLE OPTIMALE D'UN ECHANTILLON DANS L'ETUDE DES ECOSYSTEMES

L'échantillonnage est présenté comme un processus dans lequel la structure de l'échantillon implique et atteint une stabilité croissante lorsque la taille de l'échantillon augmente. La taille minimale pour laquelle la

structure de l'échantillon commence à atteindre la stabilité est proposée comme limite inférieure pour la taille d'un échantillon dans l'anayse des écosystèmes. Cette méthode et ce concept sont illustrés par deux exemples: le premier est relatif à la détermination de la taille optimale de l'échantillon dans une étude de végétation et le second, à la vérification de la valeur de ce choix.

ACKNOWLEDGEMENTS

The research described is part of a broader project under NSRC of Canada (L.O.) and CNPq of Brazil (V.D.P.) support.

REFERENCES

FEOLI E., LAGONEGRO M., ORLÓCI L. (1984). *Information Analysis of Vegetation Data.* Dr. W. Junk, bv., The Hague, 143 p.

GREIG-SMITH P. (1983). *Quantitative Plant Ecology.* 3rd ed. Blackwell, Oxford, 359 p.

JUHÁSZ-NAGY P., PODANI J. (1983). Information theory methods for the study of spatial processes and succession. *Vegetatio* 51, 129-140.

KENKEL N. C. (1984). *Boreal Vegetation of a Lacustrine Surface Sand Belt, Elk Lake, Ontario: Types and Environmental gradients.* Ph.D. Thesis. University of Western Ontario, London, Ontario, Canada, 380 p.

KENKEL N. C., JUHÁSZ-NAGY P., PODANI J. (1989). On sampling procedures in population and community ecology. *Vegetatio.* (In press.)

ORLÓCI L. (1978). *"Multivariate Analysis in Vegetation Research".* 2nd ed. Dr. W. Junk, The Hague, 451 p.

ORLÓCI L. (1988). "Community organization: recent advances in numerical methods". *Can. J. Bot.* 66, 2626-2633.

ORLÓCI L., STANEK W. (1980). "Vegetation survey of the Alaska Highway, Yukon territory: types and gradients". *Vegetatio* 41, 1-56.

ORLÓCI L. , PILLAR V. DE PATTA. (1990). "*Ecosystem surveys: When to stop sampling*". Proceedings of the 1989 International Conference and Workshop on Global Monitoring and Assessment: Preparing for the 21st Century, Venice. Fondacione G. Gini, Rome. (In press.)

PILLAR V. DE PATTA. (1988). "*Fatores de Ambiente Relacionados à Variação da Vegetação de um Campo Natural*". Tese de Mestrado, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, 164 p.

PILLAR V. DE PATTA, JACQUES A. V. A., BOLDRINI I. I. (1989a). "Fatores de ambiente relacionados à variação da vegetação de um campo natural". *Pesquisa Agropecuária Brasileira*. (In press.)

PILLAR V. DE PATTA, JACQUES A. V. A., BOLDRINI I. I. (1989b). "*Environmental related variation in a natural grassland of Rio Grande do Sul, Brazil*". Proceedings of the 16th International Grassland Congress, Nice, France. (In press.)

PODANI J. (1982). "*Spatial Processes in the Analysis of Vegetation*". Ph.D. thesis, University of Western Ontario, London, Ontario, Canada, 337 p.

PODANI J. (1984). "Spatial processes in the analysis of vegetation: theory and review". *Acta Botanica Hungarica* 30, 75-118.

POORE M. E. D. (1955). "The use of phytosociological methods in ecological investigations. II. Practical issues involved in an attempt to apply the Braun-Blanquet system". *J. Ecol.* 43, 226-244.

POORE M. E. D. (1956). "The use of phytosociological methods in ecological investigations. III. Practical applications". *J. Ecol.* 40, 28-50.

SAMPFORD M. R. (1962). "*An Introduction to Sampling Theory*". Oliver & Boyd, Edinburgh, 292 p.

SHEPARD R. N., CARROLL J. D. (1966). "*Parametric representation of nonlinear data structures*". In: P. R. KRISHNAIAH (ed.), Multivariate Analysis, pp. 561-592. Academic Press, London, 592 p.

WILDI O., ORLÓCI L. (1987). "Flexible gradient analysis: a note on ideas and an example". *COENOSES* 2, 61-65.