COMMUNITY ECOLOGY 1(1): 81-87, 2000
1585-8553 © Akadémiai Kiadó, Budapest

A comparison of sampling designs in a Hainan tropical rain forest

S. X. Yu[1,3], H. S. Y.Chan[2] and K. W.Chung[2]

[1] *Department of Biology, Zhongshan (Sunyatsen) University, Guangzhou, P.R. China.*
[2] *Department of Mathematics, City University of Hong Kong, Hong Kong, P.R. China*
[3] *Corresponding author. E-mail: lssysz@zsu.edu.cn.*

Keywords: Accuracy, Efficiency, GIS simulation, Parameter estimation, Quadrat size, Quadrat shape, Sample size, Sampling method.

Abstract. Different sampling strategies are simulated by changing quadrat size, quadrat shape, sample size and the arrangement of quadrats in a tropical rain forest of Hainan (South China). The simulation uses enumeration data of trees, and derived variables such as species richness, species importance, and species population density, to compare the efficiency of the sampling. The results verify that greater sampling efficiency is to be expected using systematic sampling than random sampling. Quadrat size has substantial influence on parameter estimation, but quadrat shape has negligible effect except when the quadrat is extremely long and narrow.

Abbreviations: GIS - Geographic Information System, DBH - Diameter at Breast Height.

1. Introduction

Sampling is information collection using only a portion of the target population. The immediate objective may be as simple as estimating one or more parameters that characterize a population, or as complex as structure identification in a collection of populations (Orlóci and Pillar 1989).

Vegetation sampling has been viewed as the first step in vegetation research (Greig-Smith 1983, Kenkel et al. 1989). It has also been considered as a process going through several cycles of successive approximation (Orlóci 1991, 1993). Many published studies address vegetation sampling with emphasis divided between theory and technique. The characteristics targeted include the spatial distribution of sampling units (quadrats), sampling unit size and shape, sample size, sample estimate stability, and sample structure stability. Studies published on these topics rarely involve tropical rain forest vegetation, which is the most structually complex and most diverse ecosystem in the world. The objective of this paper is to report results from a sampling computer simulation, applied to our study of the rain forest at the Bawangling Gibbon National Nature Reserve on Hainan Island, South China. The simulation method is GIS-based.

The utility of computer simulation in sampling or outright computerized sampling has long been recognized. Some examples focus on the determination of optimal sampling strategy (Palley and O’Regan 1967, O’Regan and Palley 1965, Arvanitis and O’Regan 1967), edge effect (Wensel and John 1969, Wensel 1975), and optimal plot number or size (O’Regan et al. 1973, Orlóci and Pillar 1989). Others have used computer simulation to measure the effect of sample size on inter-specific relationships, the scale dependence of community classifications (Podani 1987), and sample estimate or structure stability (Orlóci and Pillar 1989, Orlóci 1991).

Study area

The Bawangling forest region is located between N18^{o}50’ - N19^{o}05’ and E109^{o}05 - E109^{o}25’ in the southwestern portion of Hainan Island, about 270 km from the capital city of Hainan province, Haikou. The Gibbon Nature Reserve, situated within the eastern portion of the Bawangling forest region, was established in 1980 to protect the local population of the Hainan Gibbon (*Hylobates*

Table 1. The families and the numbers of their genera and species present in the permanent sample plots in Bawangling Nature Reserve.

Family	number of		Family	number of	
	genera	species		genera	species
Aceraceae	1	2	*Menispermacea*	2	3
Alangiaceae	1	1	*Mimosaceae*	2	4
Anacardiaceae	1	1	*Moraceae*	3	10
Annonaceae	4	7	*Myrisinaceae*	3	10
Apocynaceae	7	9	*Myrtaceae*	3	12
Aquifoliaceae	1	14	*Nyssaceae*	1	1
Araliaceae	2	4	*Olacaceae*	4	11
Burseraceae	2	2	*Palmae*	1	2
Caprifoliaceae	2	2	*Papilionaceae*	2	5
Celastraceae	1	2	*Pentaphylacace*	1	1
Cephalotaxaceae	1	1	*Pittosporaceae*	1	2
Corylaceae	1	1	*Podocarpaceae*	2	3
Daphniphyllacea	1	2	*Polygalaceae*	1	1
Dilleniaceae	1	2	*Proteaceae*	1	5
Ebenaceaea	1	5	*Rhizophoraceae*	1	1
Elaeocarpaceae	2	9	*Rhmnaceae*	1	1
Ericaceae	2	3	*Rosaceae*	4	5
Escalloniaceae	2	3	*Rubiaceae*	10	20
Euphorbiaceae	6	12	*Rutaceae*	2	3
Fagaceae	3	35	*Sabiaceae*	2	9
Flacourtiaceae	2	3	*Samydaceae*	1	1
Guttiferae	2	4	*Sapindaceae*	2	2
Hamamelidaceae	3	4	*Sapotaceae*	3	5
Hippocrateaceae	1	1	*Saptaceae*	1	1
Icacinaceae	2	3	*Sterculiaceae*	2	3
Illiciaceae	1	1	*Styracaceae*	2	2
Ixonanthaceae	1	1	*Symplocaceae*	1	22
Juglandaceae	1	2	*Theaceae*	8	9
Lauraceae	10	44	*Thymelaeaceae*	2	3
Loganiaceae	1	1	*Tiliaceae*	1	2
Magnoliaceae	4	5	*Ulmaceae*	1	3
Melastomataceae	4	5	*Vacciniaceae*	1	1
Meliaceae	2	4	Total	148	358

concolor). The reserve covers about 2,500 hectares of pristine humid tropical forest, predominantly evergreen, and about a 4,500-hectare buffer zone. Elevation varies from 400 - 1,437 m above sea-level. The climate is typically tropical with seasonal monsoons. The mean annual temperature is 24.5^{o}C, maximum temperature 38.8^{o}C, and minimum temperature 8.6^{o}C. The mean annual precipitation is 2,096.1 mm.

About 20 hectares of vegetation have been surveyed during the last decade (Yu 1989, Yu et al. 1993, 1994), in both permanent and non-permanent sample plots. Enumeration data used in this paper involved trees with a minimum of 10 cm DBH from a 350 m x 150 m (5.25 ha) permanent plot gridded into 525 quadrats each 10 m x 10 m in size. The plot is located in the *Dacrydium pierrei* + *Xanthophyllum hainanensis* - *Syzygium araicladum* vegetation type (Yu et al. 1993). Tree enumeration data include species name, tree height, diameter at breast height (DBH), crown width and depth, and tree position. The high species richness is comparable to that in the subtropics (Yu and Orlóci 1990). Table 1 lists the leading families, the number of genera and species within the 5.25 ha area. The values obtained within the plot, including species richness, species importance values and species densities, are regarded as the 'true' population parameters.

Computer simulation and data analysis

Computer simulation

With tree coordinates determined, we mapped tree distribution using the GIS software ArcView. Based on the maps, sampling can be simulated. Results are presented for sampling area size 17,500 m^2 in systematic and random sampling, and for sampling area 20,000 m^2 with varying quadrat size and shape.

Systematic vs. random sampling

The quadrat size for simulation was fixed at 10 x 10 m. Assuming that n samples will be selected from the population with elements labeled by 1,2,…,N, the generalized procedure of systematic sampling is taken by first choosing a desired sampling interval k, determining a ran-

Table 2. Size of square sampling units used in simulations.

no.	quadrat area (m^2)	side length (m)	sample size	total area (m^2)
1	100	10.00	200	20000
2	200	14.14	100	20000
3	400	20.00	50	20000
4	500	22.36	40	20000
5	1000	31.62	20	20000

dom number j and selecting the elements labeled $j, j+k, j+2k, \ldots j+(n-1)k$. In case the ratio N/n is an integer k, then j is a random number between 1 and k. Using $k=3$, the sample size is $n = 175$ selected from 525 quadrats labeled 1 to 525. We call this sampling strategy as Systematic A (SA). In random sampling, a pair of random numbers is generated by the computer and used as the lower-left corner coordinates of a quadrat. A total of 175 quadrats were chosen in this way with a total area of 17,500m^2. We call this sampling strategy as Random A (RA). The procedure was repeated three times, and three sets of random and systematic samples were generated. The data set of all 525 quadrats is called Systematic T (ST) and represents the 'true' population.

Varying quadrat size

Five sizes of square quadrats (Table 2) were used for comparing the effect of quadrat size on the analysis of vegetation structure. The process of random sampling simulation is similar to that of random sampling. They were derived from the mapped patterns. A pair of random numbers is generated by the computer and used as the lower-left corner coordinates of a specific size of quadrat. A total of 200, 100, 50, 40 and 20 quadrats were chosen respectively in this way with a total area of 20,000 m^2.

Square and rectangular quadrats

Four shapes were used for random sampling simulation: 2 m x 50 m, 3 m x 33.33 m, 4 m x 25 m and 5 m x 20 m. The size of each is 100 m^2. The total sampling area is 20,000 m^2.

Data analysis

Parameters estimated

Three parameters, usually regarded as the important features of community texture and structure, were estimated and compared with the parameters calculated from 525 10 m x10 m quadrats defining the permanent plot:

Species richness. This is simply the total number of species recorded, a criterion frequently used in the past to determine the optimal sample size in field surveys that employ the notion of minimal sampling. To be comparable with random sampling, the quadrats in systematic sampling (SA and ST) were randomly ordered before calculating the accumulated species number with increasing sampling area.

Species importance values. The importance of a species is a parameter used to indicate its role in the community, defined by Curtis (1947) as

$$IV_i = RA_i + RD_i + RF_i$$

In this, RA_i is the relative abundance of species i defined by the ratio "abundance of species i / sum of the abundance of all the species". RD_i is the relative dominance and RF_i is relative frequency of species i. These are ratios whose definition is consistent with that of relative abundance. The importance value is determined for each species in the sample, providing a vector **v** of the true species importance values.

Species density. This is the number of individuals representing species i in a sampling unit, also obtained for the separate species, providing a vector **d** of true values.

Sampling efficiency

The criterion used to compare various sampling strategies is to determine how close the sample estimate is to the true parameter T (accuracy). In case of species richness, the mean and variance are computed directly from the raw data. In the multivariate cases, i.e., species importance and density, the square of maximum error sum is the appropriate criterion. This incorporates the difference between the estimate and the true parameter summed over n dimensions:

$$SE = \sum_{i=1}^{n} d_i$$

where

$$d_i = \begin{cases} (p_i - e_i + SD_i)^2 & p_i \geq e_i \\ (p_i - e_i - SD_i)^2 & p_i < e_i \end{cases}$$

In these, p_i is the true parameter of population i, e_i is the estimate of p_i and SD_i is the standard deviation of species i estimated from sampling.

Table 3. Significance tests of sample species richness at different sampling unit sizes for systematic and random sampling. The t-value is used to test the hypothesis that the sample means in systematic and random sampling have common expectation.

Sampling area (x100m^2)	Species richness (mean ± SD)		t-value (*p<0.05; **p<0.01)
	Systematic	Random	
10	45.00 ± 1.00	44.67 ± 2.91	0.108
20	73.33 ± 3.18	70.00 ± 4.58	0.598
30	94.33 ± 2.73	85.33 ± 3.28	2.108
40	106.00 ± 2.52	99.67 ± 0.67	2.433
50	121.67 ± 3.18	113.67 ± 2.19	2.073
60	129.33 ± 4.48	122.33 ± 1.86	1.442
70	137.33 ± 4.67	128.00 ± 3.61	1.583
80	143.33 ± 3.76	134.67 ± 2.19	1.994
90	151.00 ± 3.61	141.00 ± 1.53	2.554
100	155.33 ± 4.48	144.67 ± 0.33	2.372
110	160.33 ± 2.91	149.00 ± 1.15	3.624*
120	166.67 ± 1.45	153.33 ± 1.45	6.489**
130	171.67 ± 2.40	155.33 ± 2.33	4.876**
140	174.67 ± 1.86	158.67 ± 1.45	6.788**
150	176.67 ± 1.33	160.00 ± 1.00	10.000**
160	178.00 ± 2.00	161.67 ± 0.88	7.472**
170	178.67 ± 1.86	162.33 ± 0.33	8.662**
175	178.67 ± 1.86	162.33 ± 0.33	8.662**

Table 4. The difference between the estimate and the true parameter under different sampling conditions. SD refers to the standard deviation of the mean for the difference in repeated sampling. Two parameters, importance value and species population density, are measured.

Sampling strategies		Difference (mean ± SD)	
		importance value	density
Sampling method	Systematic	62.10±23.67	9.49±0.04
	Random	81.52±30.74	9.76±0.60
Quadrat sizes	100m^2	72.97±3.32	9.23±0.56
	200m^2	90.84±25.197	5.69±0.35
	400m^2	140.86±14.529	4.64±1.77
	500m^2	158.92±41.00	3.22±0.25
	1000m^2	189.57±49.61	2.31±0.73
	2mx50m	67.56±8.92	8.65±0.12
various quadrat shape	3mx33.33m	86.06±10.63	9.11±0.06
	4mx25m	77.60±27.62	9.05±0.25
	5mx20m	104.59±16.20	8.80±0.23
	10mx10m	72.97±3.32	9.23±0.56

Table 5. Significance tests of parameter estimation accuracy for different quadrat sizes. Upper triangle elements represent the t-values of species importance value estimation. Lower triangle elements represent the t-values of species population density (*p<.05; **p<.01).

Quadrat size	100 m^2	200 m^2	400 m^2	500 m^2	1000 m^2
100 m^2		9.27**	4.29*	16.87**	13.01**
200 m^2	1.22		1.01	9.97**	7.26**
400 m^2	7.89**	2.98*		1.38	2.12
500 m^2	3.62*	2.45	0.72		2.05
1000 m^2	4.06*	3.07	1.63	0.82	

Results

Systematic vs random sampling

The species number - sampling area relationship for various sampling are shown in Figure 1a. The species richness estimate is higher in systematic sampling than in random sampling when the total sampling area exceeds 11,000 m^2 or when sample size exceeds 110. Statistical analysis indicates that there is a significant difference between the estimates (Table 3).

As could be expected from sampling theory, with the increase of sample size or sampling area, the sample estimate becomes more stable and comes closer to the true parameter. This is a condition by which to judge sample size sufficiency. The total sampling area required is about 25,000 m^2 for all species to be counted (Figure 1a). Approximately 31% or 82 species are occasional (occuring as a single individual), and an additional 12% (33 species) occurred only twice. Species richness attains a stable value as sample size increases to about 150 quadrats or sampling area 15,000 m^2 (Figure 1a and Table 3). This indicates that the lower bound of sample size is about 150 quadrats or the minimal sampling area is 15,000m^2 for enumeration of trees with a minimum of 10 cm DBH. The estimation of rare species is obviously a totally different problem.

Differences between the estimated and true value in random and systematic sampling for the parameters of species population density and species importance value are listed in Table 4. The lower difference indicates higher accuracy. The analysis suggests that there is no significant difference between random and systematic sampling in species population density estimates (t = 1.65). By contast, the importance value estimated in systematic sam-

Table 6. Significance tests of parameter estimation accuracy for different quadrat shapes. Upper triangle elements represent the t-values of species importance value estimation. Lower triangle elements represent t-values of species population density (*p<.05; **p<.01).

quadrat shape	2m×50m	3m×33.33m	4×25m	5m×20m	10m×10m
2m×50m		2.31	0.60	3.47*	0.98
3m×33.33m	6.30**		0.50	1.66	2.04
4×25m	5.42**	2.62		1.46	0.29
5m×20m	1.04	2.33	3.61*		3.31*
10m×10m	1.76	0.35	0.75	1.23	

pling is much closer to the true parameter value than in random sampling (t = 12.87, p<.01).

Optimal quadrat size

Species/area relationships for various quadrat sizes are similar (Figure 1b). Sample estimates at quadrat sizes 100 m^2, 200 m^2 and 400 m^2 have a slight higher accuracy than those estimates with the quadrat sizes of 500 m^2 and 1,000 m^2. For species importance values, estimation with small quadrat sizes is more accurate than with larger quadrat sizes (Table 4). The statistical analysis also indicates that the difference between estimates obtained at smaller quadrat sizes, 100 m^2 or 200 m^2, and larger quad-

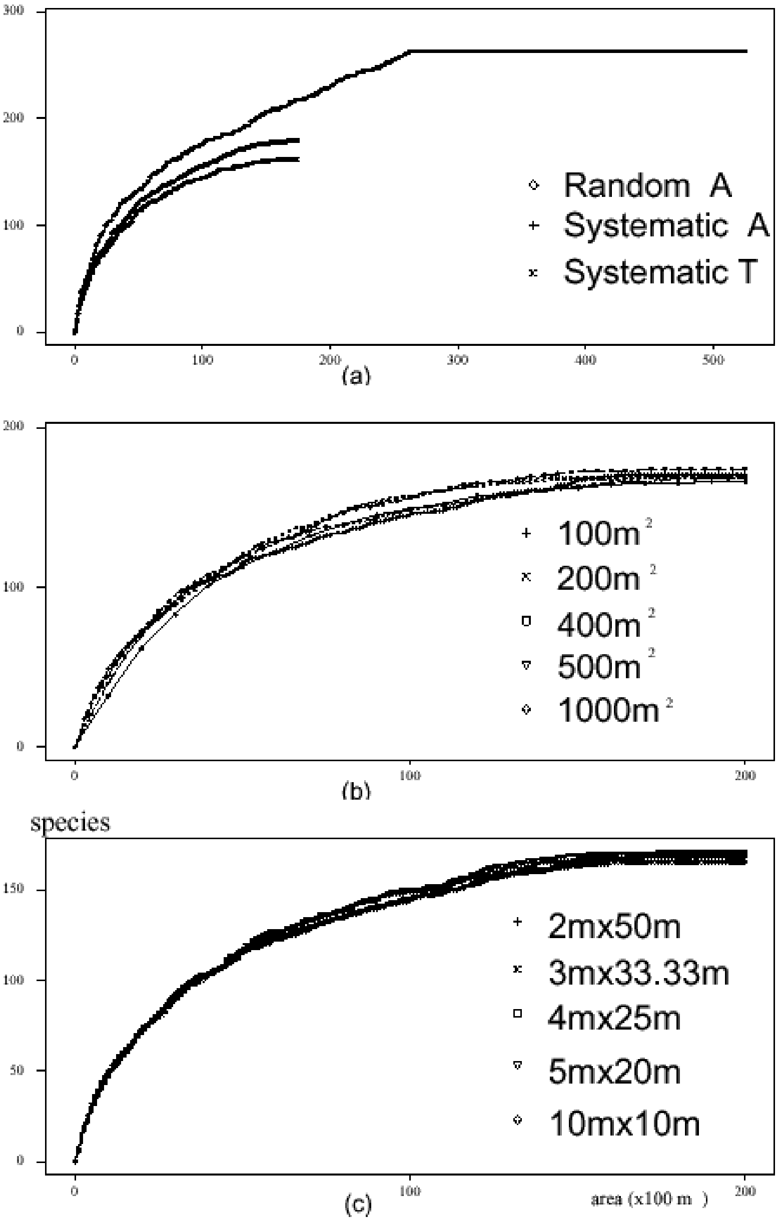

Figure 1. Species number - sampling area relationship with different sampling designs. (a) Systematic vs. random sampling; (b) various quadrat sizes; (c) various quadrat shapes.

rat sizes, 400 m^2, 500 m^2 or 1,000 m^2, is significant (Table 5). But a contrasting conclusion is indicated when species population density is estimated: estimation efficiency with small quadrat size is lower than that with larger quadrat size. This indicates that the selection of an optimal quadrat size for tropical rain forest sampling depends on the parameter to be estimated. If species population density is of interest, then the larger quadrat is preferred and will save on sampling effort in the field. However, a smaller quadrat will be a better choice for estimation of other parameters.

Optimal quadrat shape

Species richness/sampling area curves derived for various sampling unit shapes are very similar (Figure 1c). This suggests that there is no effect of quadrat shape on species richness estimation. Similar conclusion can be drawn from the estimation of species importance values and population densities (Table 4). Most of the differences among quadrat shapes are slight. For estimating species importance value, efficiency differences among quadrat shapes are not significant either, except the differences between 2 m x 50 m and 5m x 20 m, 5m x 20 m and 10 m x 10m (Table 6). For species density estimation, the estimation efficiency differences among different quadrat shapes are not significant, except the differences between 2 m x 50 m and 3 m x 33.33 m, 2m x 50 and 4 m x 25 m, 4 m x 25 m and 5 m x 20 m.

Discussion

Statistically, the objective of sampling is to achieve as high an accuracy of parameter estimation as possible for a given sample size. For vegetation research purposes, we would like to use a sample size or sampling area as small as possible whilst achieving a reasonably high sampling efficiency. The cost of sampling matters and it has to be factored into the sampling plan. For our purposes, sampling effort is a simple function of sample size and sampling unit size. The designs in our simulation experiment have focus mainly on random or systematic sampling, and on the optimal sampling unit size and shape. We believe that these are the salient aspects that need much consideration in tropical forest sampling.

In addition to random and systematic sampling, there are other related sampling methods such as restricted random sampling and stratified systematic sampling proposed in the ecological literature. From theoretical point of view, random sampling is ideal for statistical analysis. However, random sampling needs a complete "frame" (Sampford 1962), which makes pure random sampling impractical for forest ecologists. Implementation in our simulation is facilitated by the mapped tree positions in the permanent plot. With tree positions given, we could simulate random sampling with the aid of GIS software. Our results have clearly demonstrated that parameter estimation efficiency with systematic sampling is higher than with random sampling when the total sampling area is larger than 11,000 m^2.

It is recognized that an increase in sample size will result in an improved parameter estimation efficiency (Eckblad 1991, Kenkel and Podani 1991). In other words, if the sampling unit size is fixed, sample structure will attain increasing stability as the sample size increases. Species richness/sampling area relationship is a typical example for this. There are few detailed analyses from the tropical forest (Condit et al. 1996, Condit et al 1998), but our results demonstrate this clearly. For trees with a minimum of 10 cm DBH, the species richness/sampling area curve indicated that the minimal sampling area for the structural analysis of our tropical rain forest is about 15,000 m^2.

Regarding the term "quadrat" as it used in this paper, it is synonymous with "plot", both meaning a sampling unit delimited by area. The sampling unit size is decisive on the scale at which vegetation pattern should optimally be studied. Sample size is a different matter, referring to the number of sampling units in the sample. This should be as high as it is practical to make it to increase the parameter estimation efficiency. If the total sampling area were fixed, it should be prudent to use smaller quadrat size. Our simulations indicate that parameter estimation efficiency with smaller quadrats is higher than with larger quadrat size for species richness and species importance values. The situation is the opposite when species densities are estimated.

Sample units usually have square or rectangular shape in forest surveys but their circumference is the minimum for circular shapes. The accuracy gained by circular shape may however not be commensurate with the increased expense required by placing circular plots in the field. Sampling unit shape had in fact no significant effect on the structural analysis of the local vegetation.

Comparing the efficiency in parameter estimation when the sample is multivariate can be problematic. We developed a simple criterion that can be used to compare parameter estimates from different sampling strategies for the multivariate case. This method will be helpful in determining an optimal quadrat size and shape when enumerating plant communities.

Acknowledgements. This research was carried out when Shixiao Yu (S.X.Yu) was supported by a research grant to work at the Department of Mathematics, City University of Hong

Kong. The field survey was partially supported by the Natural Science Foundation of Guangdong province. The authors thank Dr. L. Orlóci for helpful editorial suggestions. Comments from two anonymous referees are also gratefully acknowledged.

References

Arvantis, L.G. and W.G. O'Regan. 1967. Computer simulation and economics efficiency in forest sampling. *Hilgardia* 38: 133-164.

Condit, R., S.P. Hubbell, J. V. Lafrankie, R. Sukumar, N. Manokaran, R.B. Foster and P.S. Ashton. 1996. Species–area and species–individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* 84: 549-562.

Condit, R., R.B. Foster, S.P. Hubbell, R. Sukumar, E.G. Leigh, N. Manokaran, Suzanne Loo de Lao, J.V.LaFrankie and P.S. Ashton. 1998. Assessing forest diversity on small plots: Calibration using species-individual curves from 50-ha plots. In: F. Dallmeier and J.A. Comiskey (eds.), *Forest Biodiversity Research, Monitoring and Modeling*. Man and the Bioshpere Series, Vol 20. The Parthenon Publishing Group. pp. 247-269.

Curtis, J.T. 1947. The palo verde forest type near Gonivaves, Haiti, and its relation to the surrounding vegetation. *Carib. Forest.* 8: 1-26.

Greig-Smith, P. 1983. *Quantitative Plant Ecology* (3 rd ed.). Blackwell, Oxford.

Eckblad, J. W. 1991. How many samples should be taken? *Bioscience* 41: 346-348.

Kenkel, N.C., P. Juhász-Nagy and J. Podani. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195-207.

Kenkel, N.C. and J. Podani. 1991. Plot size and estimation efficiency in plant community studies. *J. Veg. Sci.* 2: 539-544.

O'Regan, W.G. and M.N. Palley. 1965. A computer technique for the study of forest sampling methods. *Forest Sci.* 11: 99-114.

O'Regan, W.G., R.W. Seegrist and R.L. Hubbard. 1973. Computer simulation and vegetation sampling. *J. Wildlife Manage.* 37: 217-222.

Orlóci, L. 1991. Statistics in ecosystem survey: computer support for process-based sample stability tests and entropy/information inference. In: E. Feoli and L. Orlóci (eds.), *Computer Assisted Vegetation Analysis*, Handbook of Vegetation Science, Vol. 11. Kluwer Academic Publishers, Dordrecht. pp. 47-57.

Orlóci, L. 1993. The complexities and scenarios of ecosystem analysis. In: G. P. Patil and C. R. Rao (eds.), *Multivariate Environmental Statistics.* North Holland/Elsevier, New York. pp. 421-430.

Orlóci, L. and V. Pillar, De Patta. 1989. On sampling size optimality in ecosystem survey. *Biométrie-Praximétrie* 33:7-26.

Palley, M. N. and W.G. O'Regan. 1961. A computer technique for the study of forest sampling methods. I. Point sampling compared with line sampling. *Forest Sci.* 7: 282-294.

Podani, J. 1987. Computerized sampling in vegetation studies. *Coenoses* 2: 9-18.

Sampford, M. R. 1962. *An introduction to Sampling Theory*. Oliver & Boyd, Edinburgh.

Wensel, L. C. 1975. The treatment of boundary-line overlap in a forest sampling simulator. Hilgardia 43: 143-159.

Wensel, L. C. and H. H. John. 1969. A statistical procedure for combining different types of sampling. Forest Sci. 15: 307-317.

Yu, S. X. 1989. A primary study on the vertical distribution of species in the community space of Bawangling tropical rain forest. *Ecological Science* 8: 118-123.

Yu, S.X., H.T. Chang and B.S. Wang, 1993. Tropical montane rain forest of Bawangling nature reserve, Hainan Island. I. Permanent plot and the community types. *Ecological Science* 12: 13-18.

Yu, S.X., H.T. Chang and B.S. Wang. 1994. Tropical montane rain forest of Bawangling nature reserve, Hainan island. II. Quantitative analysis of community structure. *Ecological Science* 13: 21-31.

Yu, S. X. and L. Orlóci. 1990. On niche overlap and its measurement. *Coenoses* 5: 159-165.

Announcement

New computer program

CANOCO for WINDOWS

A new version of a highly popular software for ordination of ecological data. For more details, write to

Scientia Publishing
P. O. Box 658
H-1365 Budapest
Hungary

or visit the web site

http://ramet.elte.hu/~scientia