COENOSES 12(1):43-50, 1997
© C.E.T.A., Gorizia, Italy

SAMPLING FOR VEGETATION SURVEY: SOME PROPERTIES OF A GIS-BASED STRATIFICATION COMPARED TO OTHER STATISTICAL SAMPLING METHODS

I. Goedickemeier[1], O. Wildi[2] & F. Kienast[2]

1 Chair for Nature and Landscape Conservation, ETH, CH-8092 Zürich and Swiss Federal Institute for Forest, Snow and Landscape Research, CH-8903 Birmensdorf
2 Swiss Federal Institute for Forest, Snow and Landscape Research, Division of Landscape Ecology, CH-8903 Birmensdorf

Keywords: Mountain forest, Geographic Information System (GIS), Sampling design, Stratification, Swiss Alps, Vegetation survey.

Abstract: Within a survey of mountain forest types in the Swiss Alps, we compared three statistical sampling methods, i. e. stratified random sampling of the investigation area, systematic sampling and unrestricted random sampling. The stratification was based on a digital terrain model and performed using GIS technology. A comparison of the three sampling methods shows that stratified random sampling is an efficient method that guarantees a sample which is proportional to the extension of the environmental types. We investigated the resemblance pattern of relevés and found that it differs between purely random and all others. This leads to the conclusion that stratification introduces a bias into the true pattern. A comparison of the efficiency of the three tested sampling strategies shows that for a classification as used at the regional level (10 - 50 km^2), stratified random sampling draws an accurate picture of the small scale vegetation pattern at low sampling effort. At lower resolution, unrestricted random sampling should be applied in the present case, as it leads to a more balanced sampling of large scale vegetation patterns.

Compared to manual stratification, GIS-based stratification generally leads to a considerably higher number of strata. We found that relevés from rare attribute combinations neither tend to become outliers in the final data set nor considerably contribute to rare vegetation types. Stratification using GIS technology therefore can be recommended as an efficient technical help even if the number of resulting strata is large and rare strata have to be omitted.

Introduction

The sampling design is usually one of the most crucial factors governing the limits of results and interpretation of vegetation analysis. In general, samples may be placed in a study region in four different ways: (1) preferentially, (2) randomly, (3) systematically and (4) stratified, i.e. the study region is subdivided into compartments by some criteria and each compartment is then sampled randomly (see Gauch 1989, for an overview).

Preferential sampling is commonly used by the traditional school of Braun-Blanquet (1964). Sampling units are placed in assumingly typical, homogeneous, representative or undisturbed vegetation patches. This method lacks explicit and repeatable procedures and requires a preceding inspection of the survey area.

To avoid these disadvantages many recent investigations rely on statistical sampling to increase objectivity and to maximise the amount of biological information recorded at minimum expenditure. All types of statistical sampling methods carry some known properties, enabling the user to direct investigations towards specific objectives.

Systematic sampling has the advantage that the sampling units are easy to locate (Bourdeau 1953, Moore & Chapman 1986). But this method has an inherent tendency to over- or undersample vegetation types of extreme spatial extensions. If patch sizes vary considerably, the intersample distance is likely to be smaller than the mean diameter of the large vegetation patches and larger than the mean diameter of the small ones. Under these conditions small vegetation patches are most likely under- and large patches over-sampled. Moreover, the point pattern of systematic sampling can interfere with regular patterns occurring in the investigation area and thereby cause resonance (Finney 1949, Kershaw & Looney 1985). Another disadvantage of this method is that the position of the first point determines all the others, and if the whole grid is shifted, the results obtained may change drastically (Podani 1984 a,b).

In random sampling any part of the study area has the same probability of being sampled and the placement of a plot is completely independent of the position of those already located. The randomised arrangement ensures that the sampling is not influenced by arbitrary decisions. The data obtained in this way represent the most powerful input for statistical analyses, as they allow a quantification of the reliability of conclusions (Bourdeau 1953, Zöhrer 1980, Podani 1984a, Kershaw & Looney 1985). Compared to systematic sampling where the nearest neighbour distances are constant, random sampling has the advantage that these vary. A drawback of this method is that a large sample may be needed to represent rare types (Wildi 1986). When spatial autocorrelation between neighbouring plots is strong, systematic sampling sometimes is more efficient than a random sample of the same size (Moore & Chapman 1986).

Modifications of random and systematic sampling are attempts to overcome inherent deficiencies (see Sampford 1962, Greigh-Smith 1964, Smartt 1978, Green 1979, Caldas & White 1983, Kershaw & Looney 1985). Stratification of the area prior to random sampling to some extent prevents uneven representation of types.

Successful vegetation survey requires that the full environmental space including the extremes is covered (Margules & Stein 1989). Stratification based on environmental variables using GIS technology is expected to generate a data set which ideally represents the full environmental space. Since the technical breakthrough in the field of GIS in the late 1980s, environmental data can be stored in digital form and can be combined freely at any time to suit individual needs. Hence stratification and modelling using GIS technology has increasingly become common in vegetation surveys. Among all computerised environmental variables, climatic, topographic, soil and spectral parameters are most frequently used in vegetation surveys (Steffenson 1987, Austin & Heyligers 1989, Margules & Stein 1989, Davis & Goetz 1990, Brzeziecki et al. 1993, Neldner et al. 1995). Nevertheless, the increasing number of digitised variables generally available forces the user to carefully select the parameters based on their biological relevance. Thus, the most important decision prior to stratification is the selection of attributes to be used and the number of classes to be created. These decisions depend on the aim of the investigation, the size of the study area and the availability of appropriate computerised data sets with sufficient spatial resolution and accuracy. The more attributes are chosen for stratification and the more classes are built, the larger becomes the number of strata (Steffenson 1987). This and the usually skewed distribution of strata size lead to problems.

It is in this context that we investigated different sampling strategies for their efficiency in revealing vegetation patterns. Three criteria are used for this evaluation: the representation of the environmental space within the three data sets, the variability of species assemblage and the efficiency of the three tested sampling methods with regard to a description of the vegetation pattern under investigation.

We further inspected the role of the numerous small strata resulting from digital stratification. In this context, the question arises if relevés originating from rare attribute combinations represent 'outliers', i.e. extreme observations in the sample. Based on this we investigated whether rare strata could be omitted or, in contrary, should be sampled more intensively to yield a reliable picture of the vegetation pattern.

Material and methods

Study area

The study area covers approximately 16 km^2 on a south-facing slope in the Canton of Valais in the Swiss Alps. The forest area varies both in structure and floristic composition, ranging from xeric pine forests to montane and subalpine spruce and spruce-larch forests and to open larch forests at the timberline. The climate is dry and continental (total amount of rainfall per year about 600 mm at the bottom of the valley).

Computer systems and software

The GIS-analyses were performed on a SUN workstation applying the software ARC/INFO, versions 6 and 7. We analysed the vegetation data using the MULVA-5 statistics package for ecological data (Wildi & Orlóci 1996). The comparison of relevés is based on the similarity ratio, s, defined as

$$s = \frac{\sum xy}{\sum x^2 + \sum y^2 - \sum xy} \qquad (1)$$

where x represents the elements of one and y the same of the other relevé to be compared. The ratio has a range from 0 to 1. It has been used for classification (sum of squares clustering) as well as for nearest neighbour statistics.

Stratified random sampling using GIS technology

The GIS stratification was based on the secondary site factors elevation, slope and aspect. In a small investigation area, these parameters have nearly functional relationships to primary site factors such as temperature, soil water, radiation etc. Elevation, slope and aspect were selected based on the availability of appropriate computerised data sets with sufficient spatial resolution and accuracy. For stratification, we used the following procedure:

For each of the three attributes selected, a grid coverage of the area was derived. The elevation data originate from a national digital elevation model (DEM) with a resolution of 25 x 25 m. Aspect and slope were derived from the elevation model by using the 'aspect' and 'slope' functions in the ARC/INFO module GRID. All grid-points below 1000 m a.s.l. and above 2250 m a.s.l. (timberline) as well as the grid-points with a slope exceeding 50° were excluded. The remaining points were classified separately for each attribute (5 elevation, 3 slope and 3 aspect classes). Then the grid coverage's were transformed into polygon coverage's and overlaid to create the map of strata (see Table 1).

To restrict the sample to forest vegetation, all strata outside the forest area were excluded. Next, the total area of each stratum within the forest area was calculated. Strata with a total area below 625 m^2 were excluded. The total sample of 170 plots therefore represents a stratified random sample among the remaining strata (see Table 2).

Table 1. List of site parameters used in the stratification and class limits. DEM= digital elevation model.

Site parameter	Unit	Spatial resolution	Classes	Derivation
Elevation (DEM)	m	25 m	1000-1249 m 1250-1499 m 1500-1749 m 1750-1999 m 2000-2249 m	
Slope	deg	25 m	0-15° 15-30° 30-50°	DEM
Aspect	deg	25 m	W-N-E E-S S-W	DEM

Table 2. Sampling density in the stratified random sampling data set.

Strata area	Number of strata	Number of plots per stratum		Total number of plots	
		total sample	subsample	total sample	subsample
> 80 ha	2	10	4	20	8
20-80 ha	10	8	3	80	30
15-20 ha	6	5	2	30	12
7.5-15 ha	5	5	1	25	5
< 7.5 ha	15	1	1	15	15
Total	38			170	70

Table 3. Over-, under- and not sampled strata in the frequent and rare strata types of the systematic and random sampling data sets.

Sampling method	Total strata area	number of oversampled strata	number of undersampled strata	number of strata not sampled
Systematic sampling	> 2 ha	3	3	2
	< 2 ha	0	0	11
Unrestricted random sampling	> 2 ha	3	5	1
	< 2 ha	0	0	9

All 170 plots were used for vegetation classification. A random subset of 70 plots, served for the comparison of the three sampling methods (see Table 2).

Systematic sampling

The plots were located at regular intervals of 177 m covering the entire study area. This grid width yields the intended sample size of n=70.

Unrestricted random sampling

70 pairs of random geographical co-ordinates with a minimum distance of 25 m within the study area were chosen; thus each plot is selected independently, and each combination of co-ordinates has an equal chance of being chosen. The minimum separation of 25 m was taken to avoid overlapping plots

Field methods

At each site all species of vascular plants were recorded within a cricle of 500 square metres. Species abundance was estimated on a Braun-Blanquet (1964) cover-abundance scale. The plots were located using an aerial photograph (1:5,000) and a map (1:10,000).

Results and discussion

Representation of the environmental space

In the first step, we inspected whether the sampling schemes applied really yield different representations of the environmental space as it is needed for further comparison. Figure 1 shows the frequencies of relevés in each stratum as a function of the sampling strategy. As intended, only the data set based on stratified random sampling *a priori* covers the entire environmental space described by elevation, slope and aspect. With systematic sampling, 13 strata (34%) and with random sampling, 10 strata (26%) were not represented.

Taking the sampling density of the subset of stratified random sampling as a standard (Table 2) it is evident that systematic sampling and random sampling both lead to an under-sampling in some of the frequent strata and to an over-sampling in others (Table 3). As expected, the lack of rare strata is higher with systematic sampling than with random sampling. This is a consequence of the inherent property of the systematic sampling method to under-represented rare types.These findings of course do not yet answer the question whether the same relationships also result in the vegetation data.

Variability of species composition

To examine the variability of species composition, a nearest neighbour analysis in resemblance space was per-

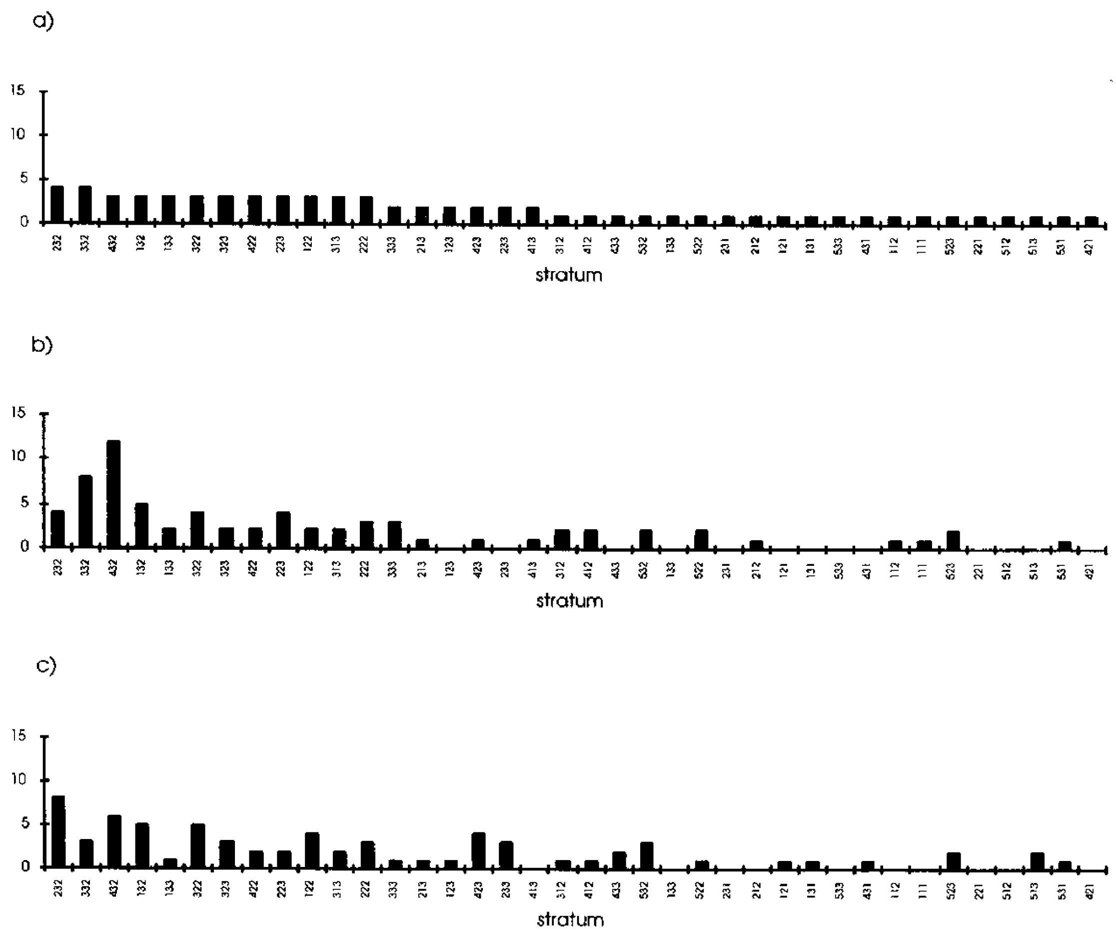

Figure 1. Number of relevés in each stratum a) stratified random sampling, b) systematic sampling and c) random sampling. The strata are coded as follows: The first digit represents elevation (1: 1000-1250 m, 2: 1250-1500 m, 3: 1500-1750 m, 4: 1750-2000 m, 5: 2000-2250 m), the second digit represents aspect (1: W-N-E, 2: E-S, 3: S-W) and the third digit represents slope (1: 0-15°, 2: 15-30°, 3: 30-50°).

formed. With this method, the similarity of each relevé to its nearest neighbour in terms of similarity is determined. The distribution of similarity values within each data set is shown in Figure 2. Isolated relevés, i.e. relevés with a low similarity with their most similar neighbours occur on the left hand side of the histogram and relevés with a high similarity to their most similar neighbours, on the right hand side.

Figure 2 reveals that the similarity distribution from stratified random sampling (Fig. 2a) is very similar to that from systematic sampling (Fig. 2b). The only obvious difference between these two distributions is a wider range of nearest neighbour similarities in the stratified random data set. We believe that the absence of a clear difference between stratified and systematic is the result of the topography and shape of our investigation area. Systematic sampling on a single slope can interfere with elevation as the main gradient. Unintendedly, the systematic sampling procedure results in a stratification of the environmental space. The effect would probably be different in a study area with a more heterogeneous topography.

Random sampling (Fig. 2c) results in lower similarity among relevés. This is surprising as the other two sampling designs are expected to suppress spatial auto-correlation. However, the opposite effect, feigning higher homogeneity in the sample than really present, appears to predominate.

The contribution of rare combinations of environmental factors (total area of stratum < 1 ha) to the resemblance pattern found in the vegetation data was assessed in two ways. Firstly an outlier analysis was performed over all 210 relevés of the three merged data sets. For each relevé the similarity to its nearest neighbour (formula 1) was determined. We considered relevés with a similarity < 0.4 as isolated. 29 isolated relevés were detected, but only 7 of them belonged to strata with rare attribute combinations.

Secondly the 210 plots were classified using sum of squares-clustering. The classification showed that relevés from strata with rare attribute combinations do not form specific vegetation types.

We therefore concluded that strata with rare attribute combinations often represent stratification artefacts due to

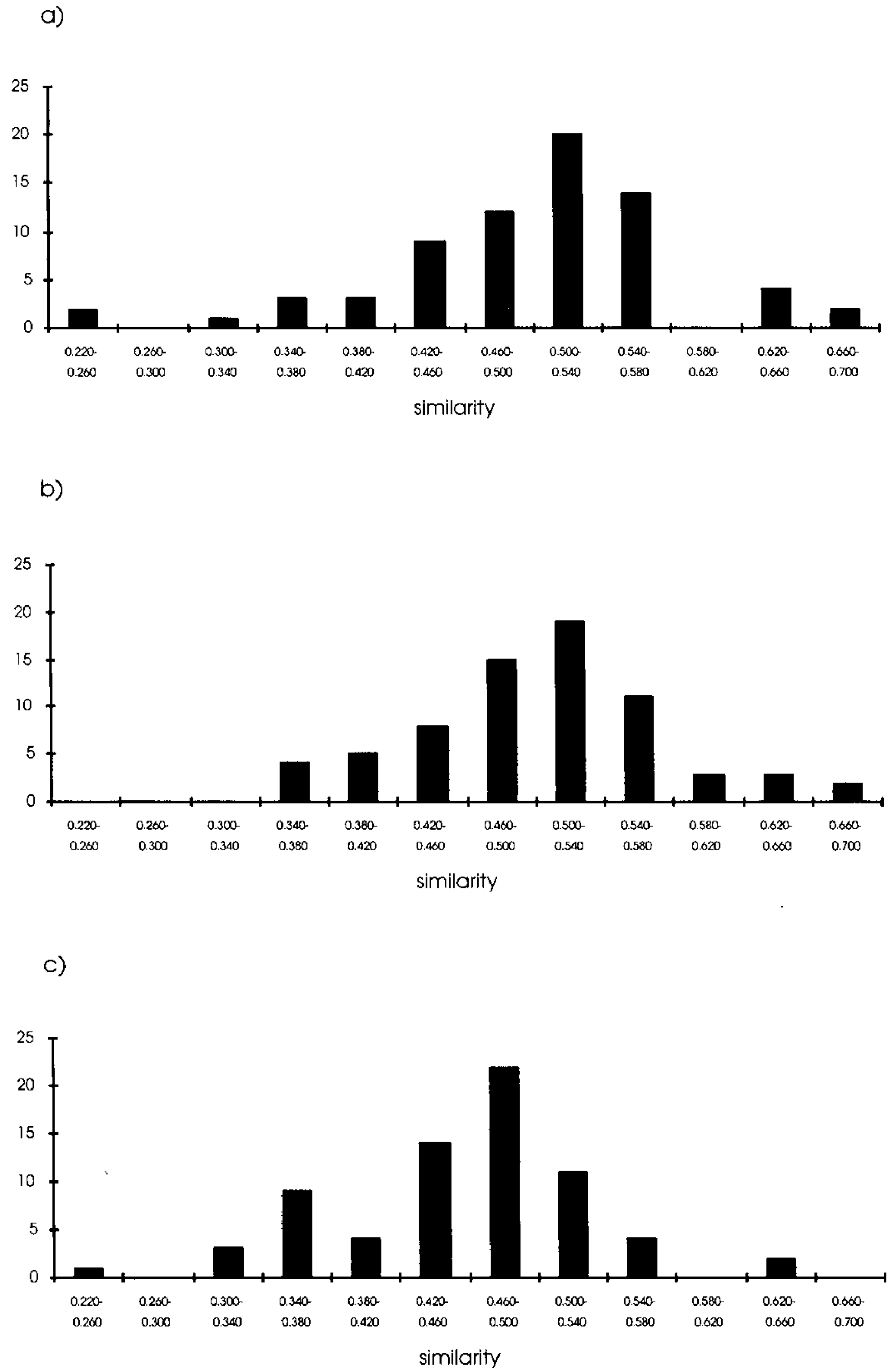

Figure 2. Histograms of nearest neighbor values, a) stratified random sampling, b) systematic sampling and c) unrestricted random sampling.

the selectional classification of the stratification variables and do not necessarily represent environmental extremes.

Representation of vegetation types

The aim of many phytosociological investigations is the conclusive recognition, description, definition and differentiation of vegetation types. The efficiency of such investigations strongly depends on the sampling design, i.e. the fewer relevés are needed, the faster and cheaper an investigation can be carried out. In order to achieve sufficient reliability of the results, the sampling design should aim at a balanced representation of all plant communities within the studied area. For that reason, it is important to know which one of the methods applied in this paper leads to a balanced representation of the vegetation at the investigation area, such that all vegetation types are described by approximately the same number of relevés. This was tested at two different hierarchical levels of vegetation classification. For this, a data set consisting of all 170 relevés from stratified random sampling was subdivided once into 3 and once into 8 vegetation types with the aid of cluster analysis. Then the relevés of the other two data sets were allocated to this classification by nearest neighbour analysis.

As Figure 3 shows, only the stratified random sampling data set yields a relatively balanced distribution of the relevés across the eight vegetation types of the lower hierarchical level. This is what can be expected in the case where the resolution of the stratification is not too different from the resolution of the final classification of the vegetation. The distribution of the relevés in the random data set is a bit more balanced than in the systematic sampling data set.

The classification at a higher hierarchical level (three vegetation units) draws a completely different picture (Fig. 4). The systematic sampling data set and the random sampling data set show a more even distribution than the stratified random sampling data set. This indicates that the stratification chosen falls short when the number of vegetation types is low.

Next, the number of relevés was reduced to 15 per data set (reduction according to the sampling design, i.e. in

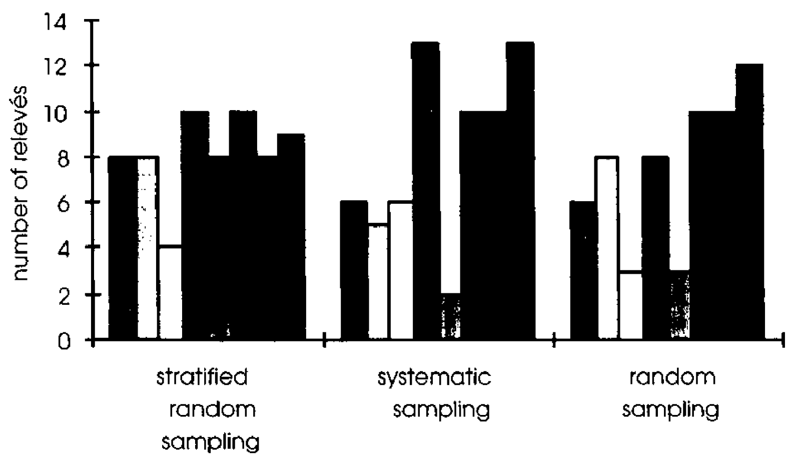

Figure 3. Number of relevés per forest type when distinguishing 8 vegetation types.

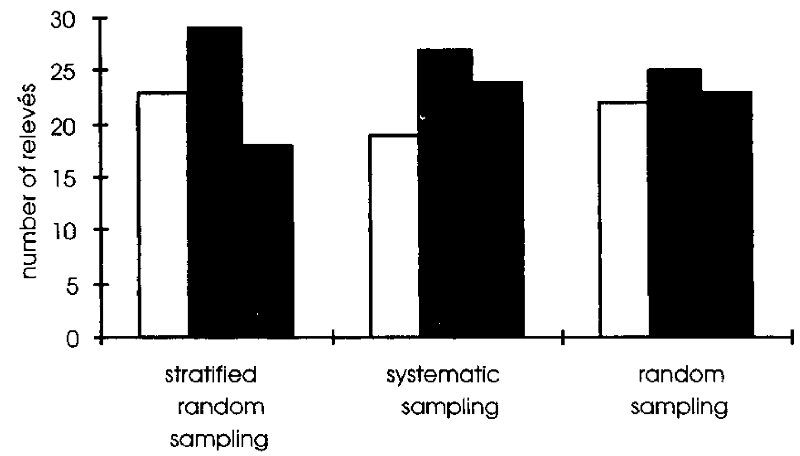

Figure 4. Number of relevés per forest type when distinguishing 3 vegetation types.

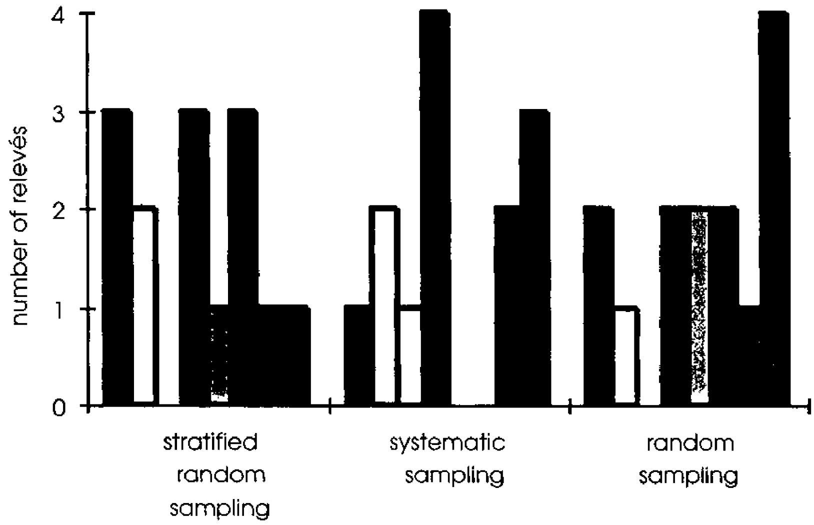

Figure 5. Number of relevés per forest type when distinguishing 8 vegetation types after a reduction of the data sets to 15 relevés.

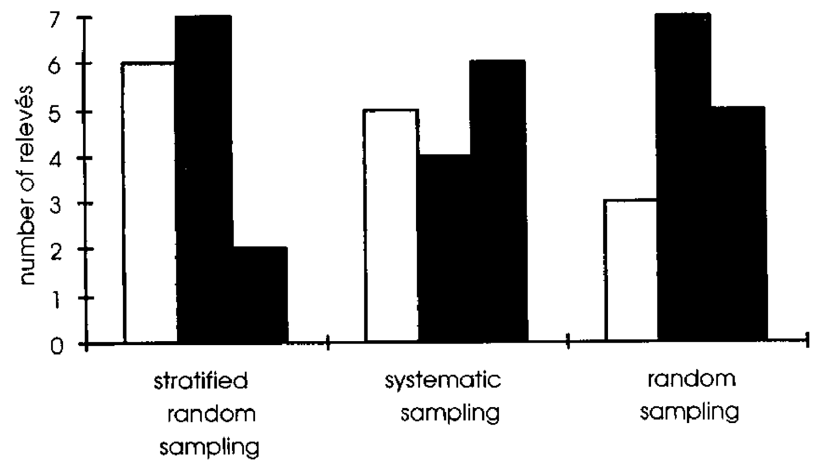

Figure 6. Number of relevés per forest type when distinguishing 3 vegetation types after a reduction of the data sets to 15 relevés.

stratified random sampling small strata were omitted, systematic sampling was reduced systematically, and random sampling randomly). As a result, the effect mentioned above is even stronger (Fig. 5) and some forest types are missing, i.e. type 3 in random and stratified random sampling and types 5 and 6 in systematic sampling. Again, the relevés of the stratified sampling data set still show the most balanced distribution over the vegetation types. On the higher hierarchical level, the classification of the reduced data sets again exhibits advantages for the systematic and the random sampling design (Fig. 6).

Stratified sampling obviously represents the environmental space, i.e. the mosaic of frequent and rare attribute combinations. Thus the high resolution classification from the stratified sampling data set directly reflects the high resolution mosaic. Systematic and random sampling do not consider the environmental space. The high resolution classification is insufficiently represented by these methods, whereas low resolution vegetation patterns, which integrate different environmental conditions, can be depicted even more accurately than in stratified sampling. This demonstrates that stratification may be preferred, but it should be used with care, as it only increases efficiency within a small range of scale.

Conclusions

In a survey of vegetation patterns in mountain forests, we stratified the investigation area based on a digital terrain model using GIS technology. To test the suitability of this sampling procedure for pattern recognition, we compared it to other statistical sampling strategies (systematic sampling and unrestricted random sampling). This comparison was done with regard to the environmental space registered, the variability of relevé resemblance and the number of vegetation types represented.

Our results show that stratified random sampling based on a digital terrain model is a suitable and efficient method for vegetation sampling. Margules & Stein (1989) demand that samples should cover the entire range of environmental variables including rare attribute combinations. Stratified random sampling based on a digital terrain model fulfils this demand. However, the gain in efficiency is paid by a loss in generality. When comparing the nearest neighbour statistics with the same of the random sample, it can be seen that the average resemblance of the relevé plots is over-estimated.

One of Green's (1979) principles for a sampling design is that each vegetation type should be represented by an equal number of samples. Stratified random sampling comes closest to this demand at a low hierarchical level. It leads to an even distribution of relevés over small scale vegetation types. However, this effect only exists when the number of strata and the number of vegetation types are high. With a high number of strata, and a low number of vegetation types, the latter tend to be unevenly represented and the advantage of the stratification vanishes.

We conclude that at scales at which the environmental conditions are reflected directly by the vegetation, stratified random sampling draws a sufficient picture of vegetation pattern at low expense. On the other hand systematic sampling and unrestricted random sampling lead to a more balanced sampling of low resolution vegetation patterns. Out of the two, unrestricted random sampling is preferable to systematic sampling for the investigation of low resolution vegetation patterns because it allows a reliable assessment of sampling error (Bourdeau 1953, Zöhrer 1980, Podani 1984a, Kershaw & Looney 1985).

A further property of stratified random sampling with GIS is that the size of the environmental strata may vary considerably. We evaluated the role of small strata in the sample in order to develop an appropriate way to handle the problem. It was found that relevés from rare attribute combinations neither tend to become outliers in the final data set nor considerably contribute to rare vegetation types. Stratification based on a digital terrain model therefore can be recommended even if the number of strata is large and strata with rare attribute combinations have to be omitted.

References

Austin, M.P. & P.C. Heyligers. 1989. Vegetation survey design for conservation: gradsect sampling of forests in north-east New South Wales. Biol. Conserv. 50: 13 - 32.

Bourdeau, P.F. 1953. A test of random versus systematic ecological sampling. Ecology 34: 499 - 512.

Braun-Blanquet, J. 1964. Pflanzensoziologie. Springer Verlag, Wien-New York.

Brzeziecki, B., F. Kienast & O. Wildi. 1993. A simulated map of the potential natural forest vegetation of Switzerland. J. Vegetation Sci. 4: 499 - 508.

Caldas, F.B. & R.J. White. 1983. Sampling for vegetation survey: a generalized flexible model for sample location. J. Biogeogr. 10: 11-28.

Davis, F.W. & S. Goetz. 1990. Modelling vegetation pattern using digital terrain data. Landscape Ecology 4, 1: 69-80.

Finney, D.J. 1986. An example of periodic variation in forest sampling. Forestry 23: 96 - 111.

Gauch, H. G. 1989. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York.

Greigh-Smith P. 1964. Quantitative plant ecology. Butterworth, London.

Kershaw, K.A. & J.H.H. Looney. 1985. Quantitative and dynamic plant ecology. 3rd ed. Edward Arnold, Victoria, Baltimore.

Margules, C.R. & J.L. Stein. 1989. Pattern in the distributions of species and the selection of nature reserves: an example from *Eucalyptus* forests in south-eastern New South Wales. Biol. Conserv. 50: 219-238.

Moore, P.D. & S.B. Chapman. 1986. Methods in plant ecology. 2. ed. Blackwell, Oxford.

Neldner, V.J., D.C. Crossley & M. Cofinas. 1995. Using geographic information systems (GIS) to determine the adequacy of sampling in vegetation surveys. Biol. Conserv. 73: 1 - 17.

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.

Sampford, M.R. 1962. An introduction to sampling theory. Oliver and Boyd, Edinburgh.

Smartt, P.F.M. 1978. Sampling for vegetation survey: a flexible systematic model for sample location. J. Biogeogr. 5: 43 - 56.

Smartt, P.F.N. & J.E.A. Grainger. 1974. Sampling for vegetation survey: some aspects of the behaviour of unrestricted, restricted, and stratified techniques. J. Biogeogr. 1: 193 - 206.

Steffenson, J.R. 1987. Application of a Geographic Information System to design a vegetation resource inventory. GIS 87 ' .. into the hands of the decision maker'. American Congress on Surveying and Mapping, 26 October, 1987. American Society for Photogrammetry and Remote Sensing, San Francisco. 431 - 439.

Wildi, O. 1986. Neue Methoden der Vegetationskartierung. Vierteljahresschr. Nat. forsch. Ges. Zür. 131 (3): 183 -189.

Wildi, O. & L. Orlóci. 1996. Numerical exploration of community patterns. A guide to the use of MULVA-5. 2nd edition. SPB Academic Publishing bv, Amsterdam.

Zöhrer, F. 1980. Forstinventur. Parey, Hamburg, Berlin.