COENOSES 5 (1): 51-60, 1990

SAMPLING WITH MULTIPLE OBJECTIVES AND THE ROLE OF SPATIAL AUTOCORRELATION[1]

O. Wildi, Swiss Federal Institute for Forest, Snow and Landscape Research, Landscape Division, CH-8903 Birmensdorf, Switzerland

Keywords: Alpine ecosystem, Sampling design, Spatial autocorrelation, Simulation, Switzerland

Abstract. The handling of mixed type multivariate data sets is discussed on an example from the "Man and Biosphere" project, Davos, Switzerland. The data comprise variables of different spatial resolution, aggregation and reliability. Correlograms are computed for data subsets describing different landscape features on differing scales. Some variables show spatial independence, others exhibit correlation among adjacent sampling localities. Periodicity is detected in the distribution pattern of some animal species, and soil types form coenoclines. While some results derived by previous modelling reflect local patterns, others suggest wide ranging relevance. Optimum sampling intensity therefore should not only depend on the aims of the study, but be also influenced by the nature of the variables. For investigations with multiple objectives, the simultaneous use of a combination of sampling designs is suggested. Quadrat size, grid width, and even investigation areas may vary. The commensurability of the designs can be achieved by simultaneously running the operations at the different aggregation levels for the relevant variables.

Introduction

The tools of conventional statistics encourage experimental designs, where different treatments are studied on independent elements of a sample. Environmental investigations, however, take place in natural space and are often time dependent. The sample elements are multivariate measurements within points or quadrats. They are likely to interact through the exchange of energy, matter and information in time/space. Variation then becomes a function of neighborhood distance, spatial autocorrelation occurs (Cliff and Ord 1981, Sokal 1986), and the independence of quadrats, the sampling units, is no longer granted. A sample can exhibit a spatial pattern deranged by processes (Upton and Fingleton 1985, pg. 27), resulting from an impact. As Green (1979) points out, the study of a changing pattern is often the only method to detect an as yet unknown or even non-measurable impact.

An efficient sampling design is aimed at facilitating the recognition of existing patterns, while keeping the effort minimal. Therefore, optimization of a study requires explicit definition of objectives and recognition that we are dealing with many variables simultaneously, possibly of different type, order of magnitude and variability, and different spatial behaviour. It follows that by optimizing a design for measuring the association of tree species, for example, we may fail to find association among animals although we finally wish to detect the relationships among trees and animals. Processing multiple observations is therefore essential in a study with multiple objectives.

The present paper describes an ecological investigation -- with variables of entirely different kinds, such as altitude, soil- and vegetation type, abundance of mammals and birds as well as land use (Wildi and Ewald 1986) -- which demonstrated the sensitivity of the natural system's state on environmental management. In implementing the investigation, the first step involved a study of the natural system (*i.e.*, the system's state in space and time) from maps and data by remote sensing and the determination of functional relationships among the elements. The subsequent step entailed the construction of a model and simulatation of potential spatial and temporal changes (Binz and Wildi 1988). Where maps or photographs are the sources of information, all elements are enumerated, so that the usual statistical considerations of sampling do not apply. However, for the construction of a simplified model, sampling the raw data may be required. Since the study area is the same for all variables, their reliability depends on type and pattern. Although the model as it stands shall not be discussed in this paper, the reliability of the results is considered. The different types of variables involved have to be investigated for this. Primarily, it is of interest to determine which variables are spatially independent at varying grid width, and if linearity or non-linearity dominates the relationships. Also, the presence of regular patterns, like oscillations, are of interest, since they can indicate that the sampling design falls short on being efficient (Greig-Smith 1983) and that there are limitations to the generalization of the results.

1 Lecture presented at the 2nd C.E.T.A. International Workshop on Mathematical Community Ecology, November 19-25, 1988, Gorizia, Italy.

Whenever spatial or temporal interactions are sought, correlograms play an important role (Dutter 1985, Upton and Fingleton 1985). They measure correlation among spaces (physical, environmental) in a stepwise fashion, considering small distances (or large similarities) first and large distances (or small similarities) at the end (Sokal 1986). Correlation, therefore, becomes a function of distance and it is usually expressed as a graph (Legendre and Fortin 1989). The type of space is unimportant, since the principle is not only applicable in spatial and temporal, but also in biological and ecological contexts (*e.g.*, Feoli and Ganis 1986).

The data

The analysis presented in the sequel is performed on data from the "Man and Biosphere" project (MaB-6 mountain programme), Davos, Switzerland (Wildi and Ewald 1986). Its aim was to predict the joint effects of economical activities, tourism and land uses on the state of the natural system of a test site situated around Davos at 1500-3200 m a.s.l. The field investigations covered different disciplines, such as geomorphology, soil science, climatology, geobotany, forestry, game biology, landscape history and landscape management. All space related information was stored in a Geographical Information System (GIS) for further processing (Fig. 1). The internal organization is related to the sampling design as follows.

Firstly, the investigation area is delineated in order to define the range within which the model shall operate. Fig. 2 shows the topographical situation and the arbitrarily established boundaries. As shown in Fig. 1, the data base can be interpreted as a digital image with several information layers, where information is stored for each element called pixel (Besse *et al.* 1982). A pixel represents an element of a 50m square grid. Operations take place at this level with 37,457 pixels for the entire 100 km^2 investigation area. If a pixel is interpreted as a basic element of information, then the entire set of pixels (*i.e.*, the raster image) can be subjected to statistical sampling at various intensities (Fig. 3). At a grid width of 250 m, the elements are reduced to 1599. Taking the 50 m pixels at the intersections of the km^2 coordinates, results in a systematic sample and cuts the size to 94 elements (heavy lines in Fig. 4).

Multiple objectives require the processing of different types of data. While some, such as altitude, soil depth, radiation and agricultural yield are continuous and metric, others, *e.g.*, soil quality, are ordinal, and a majority (soil type, vegetation type, land use, etc.) is nominal. However, this is no serious problem, as methods to process mixed data are readily available (Gower 1971). In many cases, a set of mixed data can be converted to one type (Green 1979, p. 81 ff). The change of spatial patterns among the variables are more difficult to deal with. Different cases can be distinguished.

Some of the variables are related to the altitudinal pattern of the investigation area. An example is shown in Fig. 5 with the belt of woodland, covering the steep slopes around the valley bottoms and ending at the timberline. Accordingly, the abundance patterns of woodland inhabiting animals is similar and also dominated by the pattern of the two main valleys. Then, there are patchy structures which occur at any altitude. Ski tourism (slopes, tracks) is a typical case (Fig. 6). While this map has a rather schematic appearence, agricultural yield, a quantitative and continuous variable, offers very detailed spatial information (Fig. 7). Finally, most variables derived from the simulation have much higher resolution than the mapped original. The simulated distribution of chamois is typical (Fig. 8) in that there are patches as well as many isolated pixels where the animal should occur. This results from the structure of the simulation model explained next.

Modelling the interaction of variables

The different types of variables could be considered separately. But in the present case, they are assumed

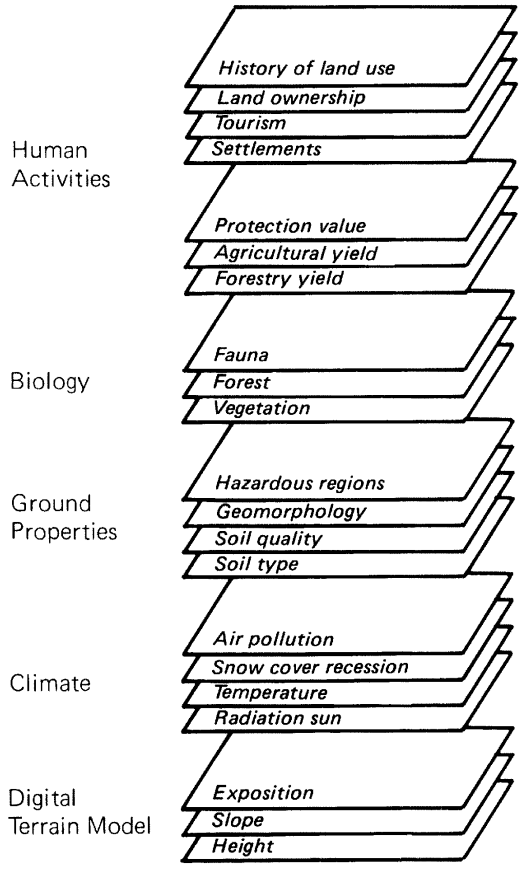


Fig. 1. Information layers of the Geographical Information System MaB-Davos (from Keller and Seidel 1984).

to interact within the simulation model. This is based on a static approach as shown in the scheme of Fig. 9. It consists of a number of submodels, which stepwise reproduce the state of the system. Parameters such as the shape of the landscape (the digital terrain model DTM, in- and outside the investigation site), geomorphology and radiation are assumed to be constant. Combined with actual or assumed landuse, a first submodel simulates soil type. This is an entirely expert-based construct, including the knowledge on how the evolution of soil is functioning. The simulated soil type is then accepted as an additional source of information for the next submodel that simulates the development of the vegetation types from a matrix of conditional state probabilities as found in the mapped information (the original data base). More detailed explanations on this are given by Fischer (1990) in this volume. Natural hazards are calculated next. Figure 10 gives an example for avalanche hazard in the same landscape with and without woodland. Mechanical functions describe the flow of snow down slopes (Binz and Wildi 1988) as against multiple regression, which is the bases for simulation of the abundance of different animal species. Other variables are achieved by weighting existing information. Agricultural yield stems from the vegetation map and uses a table containing the specific average annual production for each vegetation type.

As the physical space is described with variable precision depending on type and source of the variables, the reliability of the results is analysed in some detail.

The method

The method of multivariate spatial correlograms has been introduced by Sokal (1986) and discussed by Legendre and Fortin (1989). Basically, two spaces are distinguished, the physical and the ecological. The physical space is described by the x and y coordinates

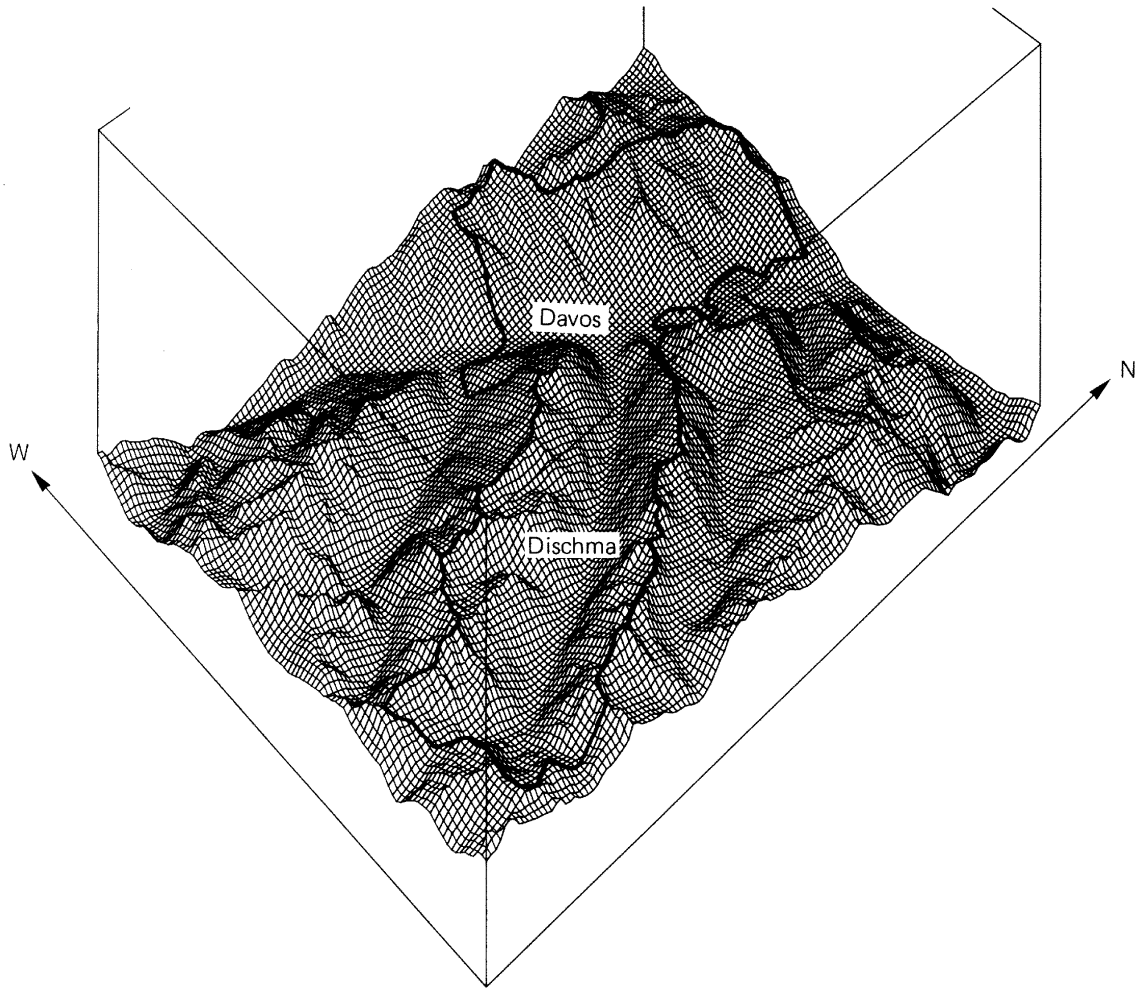


Fig. 2. Topographical conditions and boundaries of the investigation area.

of the sampling locations. The shape of the space is defined by the triangular matrix, including all combinations of distances between the points. The most obvious measure for any distance d between location i an j is the Euclidean,

$$d_{(i,j)}=\sqrt{(x_i-x_j)^2 + (y_i-y_j)^2}$$

The ecological space is defined by a matrix of multivariate resemblance measures $s_{(i,j)}$ among biological and physical records at the same places (Orlóci 1978). Moran's I (Silvertown 1980) is then used to correlate the spaces,

$$I = \frac{\sum_i d_i S_i - \frac{1}{k} \sum_i d_i \sum_i S_i}{\sqrt{[(\sum_i d_i^2 - \frac{1}{k}(\sum_i d_i)^2]\ [\sum_i d_i^2 - \frac{1}{k}(\sum_i S_i)^2]}}$$

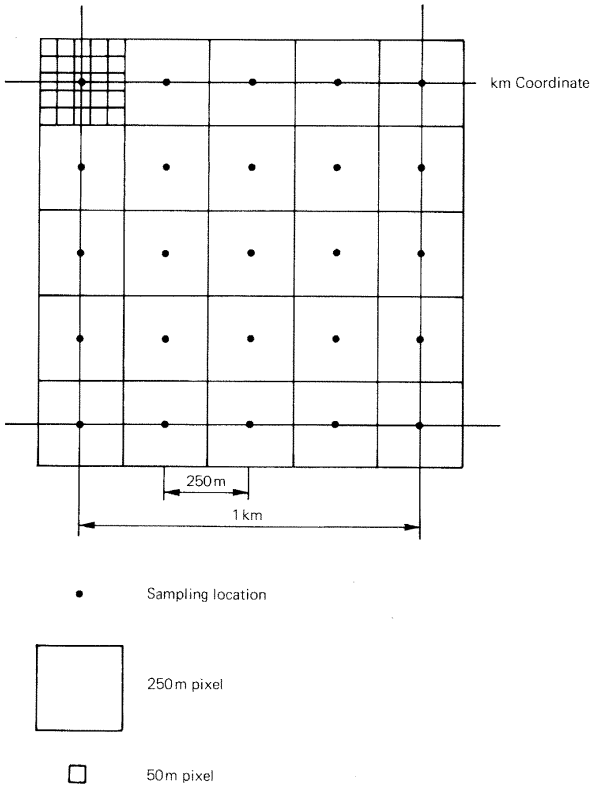


Fig. 3. The nesting of systematic sampling strategies within the data base, consisting of 50 m square pixels.

$i=1, \ldots, k$ and $k=(n(n-1))/2$. In these, n is the number of data points. As pointed out by Sokal (1986), Moran's I is structually similar to a product-moment correlation coefficient, but its upper bound is usually less than unity. To identify significant I-values, the Mantel-test (Mantel 1967) would have to be applied. In the present context, however, interpretation relies on the shape of a curve rather than a single value. This correlogram expresses the change of correlation as a function of distance classes. The original distance matrix is replaced by a weight matrix, following Sokal (1986): "For each distance class the weight matrix is given as a binary weight matrix, where the weight of 1 between a pair of localities indicates that the pair falls in this distance class and 0 indicates that it does not. Using the binary weight matrix for each distance class, one computes the corresponding spatial autocorrelation coefficients and plots them against the geographic distance implied by the distance class."

This type of correlogram is direction-free, which means that all distances have equal weight irrespective of their directions. A directed correlogram is achieved, when projecting the distances down onto a

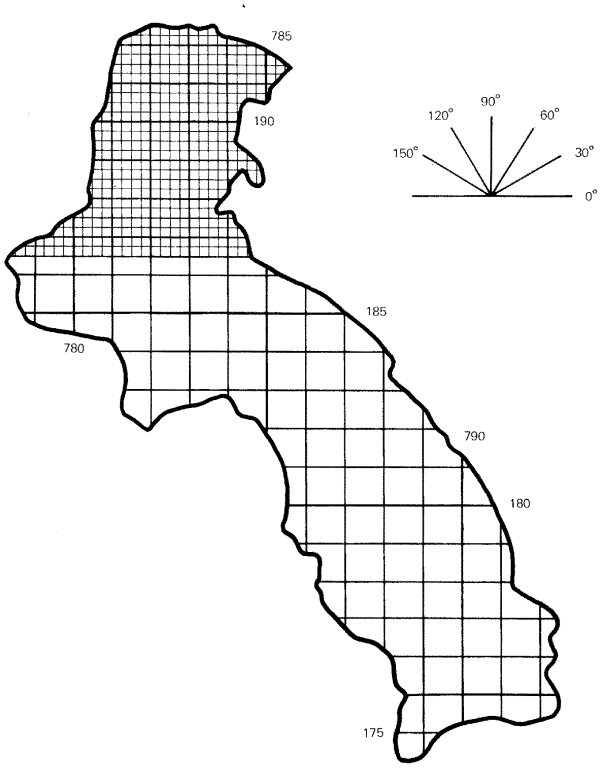


Fig. 4. Two systematic sampling designs used in the analyses. One series of data sets includes sampling locations at 1 km grid width (heavy lines). The other covers only the northern part, but operates within a 250 m square grid. The semicircle indicates directions considered in the analysis of spatial autocorrelation.

Table 1. The three sets of variables analyzed.

Variable set	Number of variables	Type of variables	Transformations	Spatial distribution pattern
"soil" (soil types)	15	nominal (presence-absence)	none	clumped, patchy
"cpar" (continuous parameters)	5	metric and ordinal	normalized	continuous
"animals" (animal species)	12	nominal (presence-absence)	none	dispersed, zonal

specified line with direction φ. For this, the direction of the connection between any two points, expressed by the angle α is determined. The deviation from the pre-conceived direction of the correlogram is given as

$$\beta = \varphi - \alpha$$

and the new distance corresponds with the projection

$$d' = d \cos \beta$$

An alternative would be to only consider connections between points that lay within a limited range of angles.

The physical space does not need to be two-dimensional. It may be extended to three or more dimensions, or even limited to one. In the case of one dimension, the distance reduces to

$$d_{(i,j)} = \text{abs}\ (x_i - x_j)$$

and the computation of directions becomes meaningless.

Fig. 5. The belt of woodland within the investigation area (from Wildi and Ewald 1986).

Fig. 6. Winter tourism. Dark areas are ski tracks, grey patches are generally accessible (from Wildi and Ewald 1986).

Results

In order to analyse the multivariate interactions of the different variables with the physical space, three groups are distinguished. They vary in type (metric, ordinal, nominal) and spatial aggregation (clumped, dispersed, continuous, Table 1). In the set of 15 soil types ("soil"), the most and the second most abundant is recorded in any location. The set of five continuous parameters ("cpar") includes height, slope, sunshine in July (h/month), agricultural yield, and soil depth. They are normalized to compensate for different scales. All variables in the set of animal species ("animals") exhibit zonal patterns. Unlike the soil types, quadrats where animals were observed are rare and scattered. No more than up to three species are usually present in any quadrat.

In order to investigate the different spatial aggregations of the quadrats, two sampling designs are tested with each set of variables. The first includes all 50 m sq. quadrats at the intersections of the 1 km grid shown in Fig. 4. The resulting sample of 94 plots covers the entire investigation area. The second sample covers the uppermost (*i.e.*, northern part), including one single south-east exposed main slope (Fig. 4). The grid width is only 250 m and the sample size is 400.

The undirected and the directed correlograms of all six data sets are shown in Tables 2 and 3. Four phenomena can be distinguished. First, at 1 km grid width (Tab. 2), the "soil" correlograms show the existence of gradients with limited range. Plots in close neigborhood are positively correlated. When increasing distances are considered, correlation decreases and eventually becomes negative. At even larger distances, correlation vanishes. The soil map seems to present a number of coenoclines of limited size, roughly 10 km in diameter. The pattern is even more pronounced in some of the directed correlograms, *e.g.*, at 60 degrees, where the investigation site is rather narrow. On the sample covering only one main slope (Table 3), the expected coenocline emerges. A second phenomenon, high and almost constant correlation at distances close and up to 500 m is due to the low resolution of the soil map, which predominantly consists of patches that size.

A third phenomenon has to do with the set of continuous parameters, "cpar". No pattern of autocorrelation is observed. Neither a set of coenoclines, nor periodicity or regular patches can be detected in the multivariate structure (Table 2). Yet at the resolution of 250 m (Table 3), oscillations which are not easy to

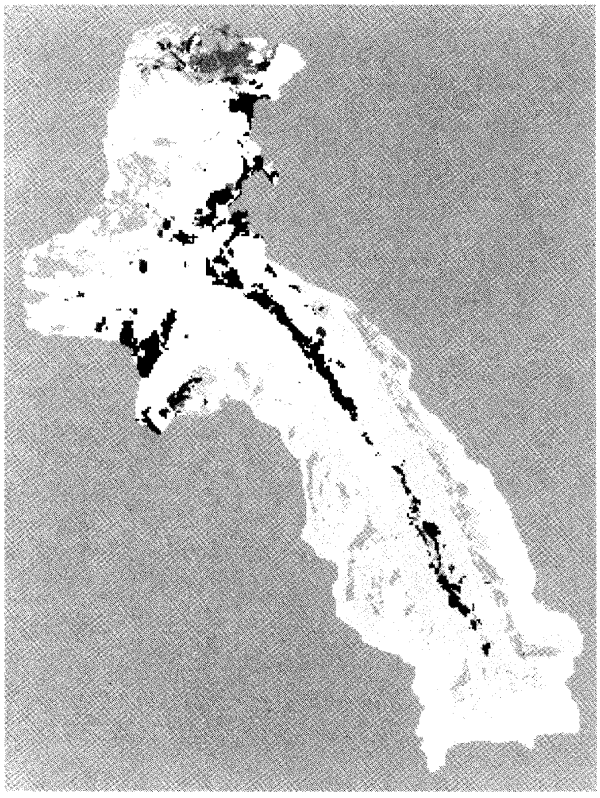


Fig. 7. Agricultural yield. Black indicates maximum yield (up to 70 dt/ha*a dry matter), white signifies zero yield (from Wildi and Ewald 1986).

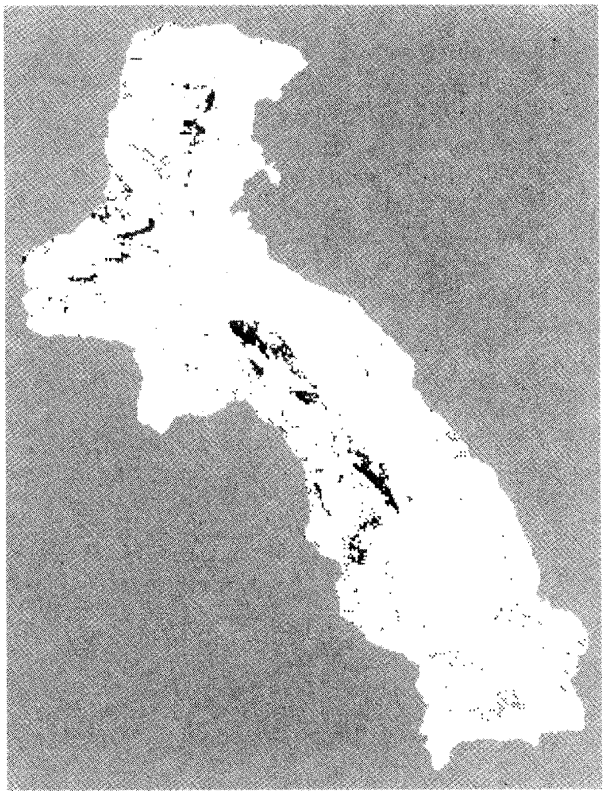


Fig. 8. Simulated abundance of chamois (*Rupicarpa rupicarpa*) in the winter (from Wildi and Ewald 1986).

Table 2. Correlograms from the 1 km square grid shown in Fig. 4. Sample size is 94. Distance classes have a width of 1 km. The variable sets are characterized in Table 1.

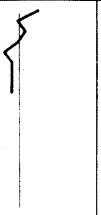
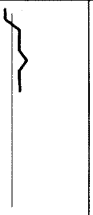
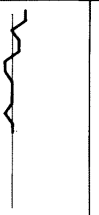
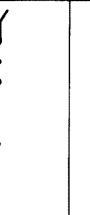
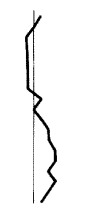
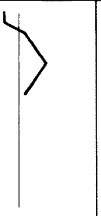
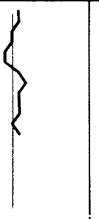
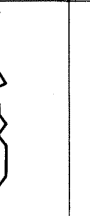
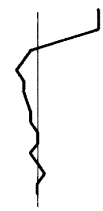
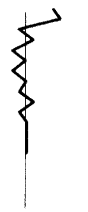
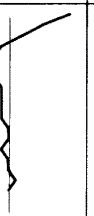
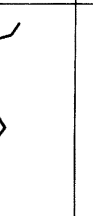
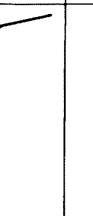
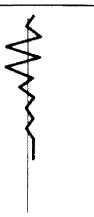
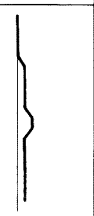
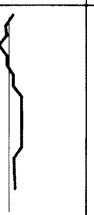
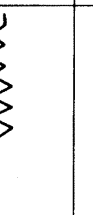
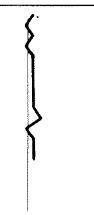
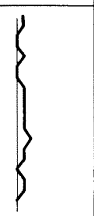
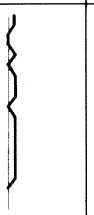
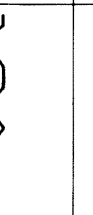
parameter set	Distance class	undirected	0 °	30 °	60 °	90 °	120 °	150 °
		-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2
„soil"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							
„cpar"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							
„animals"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							

Table 3. Correlograms from the 500 m square grid shown in Fig. 4. Sample size is 400. Distance classes have a width of 500 m. The variable sets are characterized in Table 1.

parameter set	Distance class	undirected	0 °	30 °	60 °	90 °	120 °	150 °
		-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2	-0,2 0 0,2
„soil"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							
„cpar"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							
„animals"	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20							

explain occur in some directed correlograms.

A fourth phenomenon, periodic patterns in the correlograms, is found in the set of animal species, "animals". Correlation is near zero at low distances, decreases, then increases again to produce an s-shaped curve. This structure results from the morphology of the investigation area, where three large slopes dominate. The periodicity vanishes in the 250 m grid (Table 3). The coenocline to be expected due to the dominating slope does not occur. It appears that the quadrats where animals are observed are too much scattered to reveal the underlying pattern.

Conclusions

Differences in the spatial resolution of the information limit the analyses in applications with multiple objectives. To some extent, this can be controlled by the coordination of the data collection for the various disciplines. In other cases, the lack of sharpness is due to the nature of the variables. Since animals move, their position during a season has to be described by a pattern of probabilities. This pattern then reveals a typical structure, *i.e.*, a zonation pertaining to the altitudinal shape of the investigation area. The latter is just about large enough to unfold one typical transect across one valley. Whether this structure repeats itself or not is unknown. For this to be tested, a much larger area would have to be investigated. This, on the other hand, conflicts with the effort required for taking samples to identify the soil types.

How then can the sampling efficiency be increased? One solution would be to conduct investigations in one dimension rather than in a two-dimensional grid. The use of transects is in fact widely discussed in the literature (*c.f.* Whittaker 1978, Gauch 1982, Greig-Smith 1983) and a large number of methods to analyse such data are known (Upton and Fingleton 1985). As long as the objective of the investigation is to analyse one specific, and hence arbitrarily chosen transect, this approach is most convenient. In a two-dimensional case, however, the elements of a sample may be transects and the entire sample may describe the variations among these. Unlike in the case of a single transect, the method of computing correlograms allows to determine the direction of the most pronounced coenocline. An example is shown in the first row of graphs in Table 2, where the gradiental nature of the soil types is most striking at an angle of 60 degrees. Spatial correlograms therefore represent a more general tool to gradient analysis than transect investigations.

The problem already arises when using sampling quadrats. Interactions between vegetation and soil are investigated most efficiently at the resolution chosen for the definition of vegetation types. But for the interaction of vegetation and animals, larger areas have to be

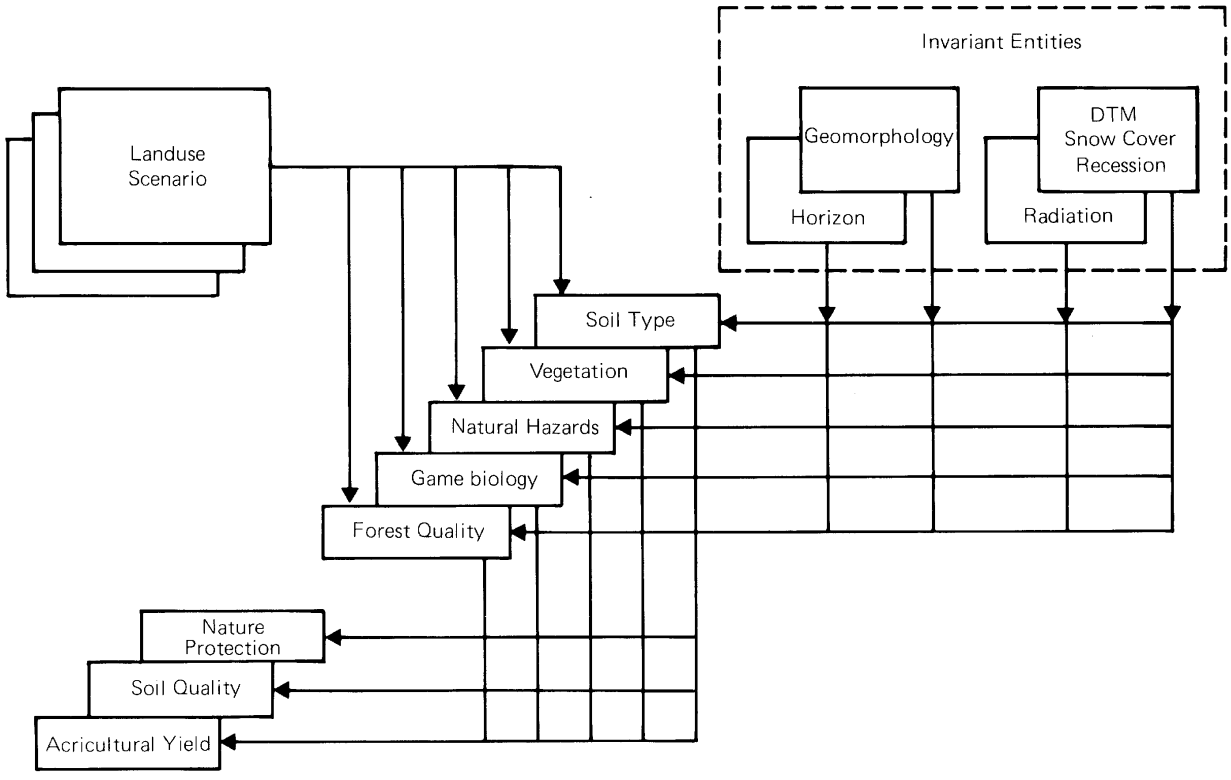


Fig. 9. Simplified schema of the model structure.

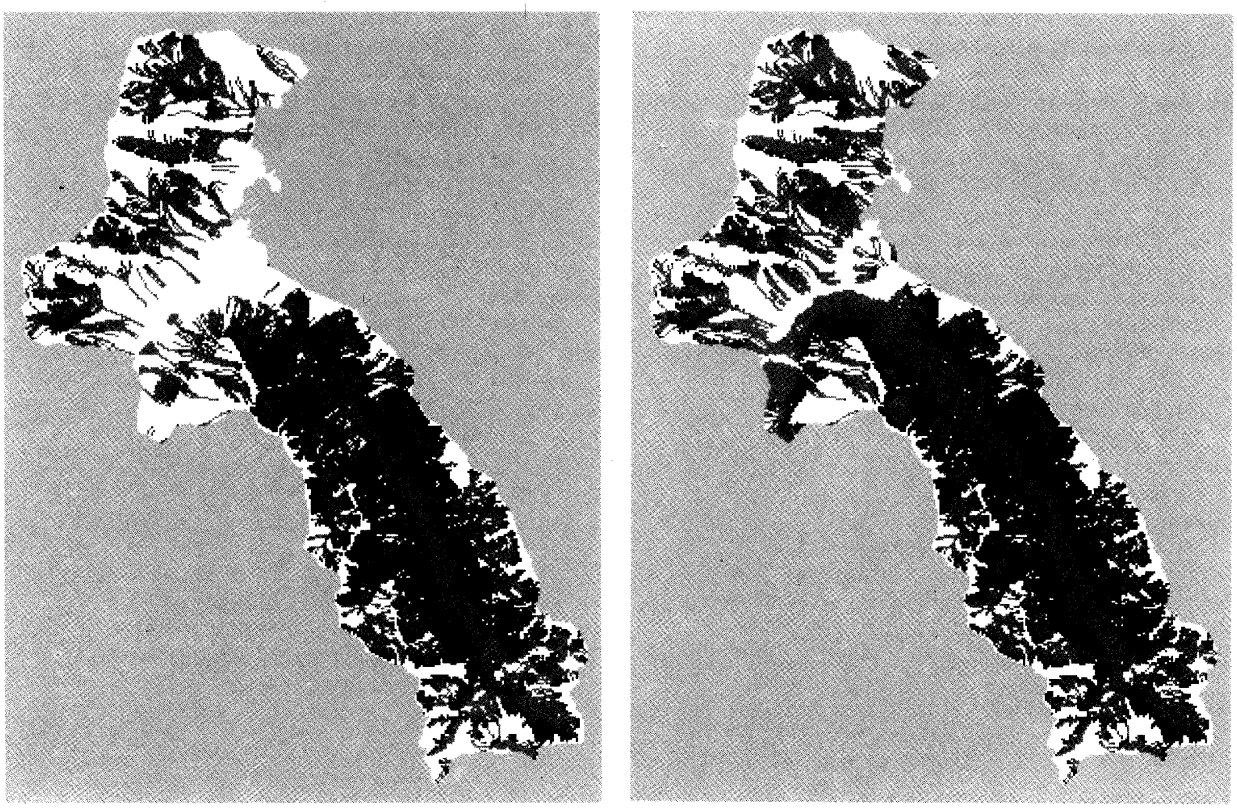


Fig. 10. Simulated avalanche hazard with woodland (left) and without (right) (from Wildi and Ewald 1986).

considered which are totally non-homogeneous with respect to vegetation. The autocorrelation pattern in the data set of animals (Tables 2, 3) also suggests the use of a wider sampling grid to achieve independent sampling elements. Unlike for vegetation analysis, a larger investigation area should be considered to reveal variation in the zonal pattern.

The conclusions drawn from the study, involving this multiple objective investigation of a complex natural systems, are as follows: First, a set of thematical maps is a rather unefficient source of information, as it presents heterogeneous information at the same scale. Second, an investigation is only efficient when sampling the investigation area (or available maps). The spacing used between the quadrats can manage the spatial autocorrelation and the spatial resolution problem, pertaining to the different types of variables. Third, to achieve commensurability among different types of variables, alternative spatial aggregations of some of the variable sets have to be considered. For example, it is best to use a detailed classification of vegetation to investigate its relation to soil type, but an aggregated system is required if animals are considered. Fourth, the investigation area may have to be varied to unfold the patterns for the different variables. As an example, a wide but also large sampling grid would be needed to analyse interactions among vegetation and animals. This means that for each type of interaction sought, a specific sampling design would have to be developed. An approach to analyse such a framework of designs must then be found.

These considerations represent approximations to an as yet unresolved statistical problem. As Sokal (1986) mentions, the determination of spatial autocorrelation relies on the assumption that the variables mapped onto the set of localities are homogeneous over this set. But since these variables are often autocorrelated, the assumptions of many tests used are often not met. Ecological work then relies on approximations.

Acknowledgement. This project was sponsered by the Swiss National Research Foundation. The author expresses his thanks to Mrs. M. Frech for helpul suggestions.

REFERENCES

BESSE, L., K. SEIDEL and O. KÜBLER. 1982. A Large Scale Multipurpose Interactive Image Processing Facility at ETH-Zurich. In: J.L. Mannos (ed.), *Design of Digital Image Processing Systems.* Proc. SPIE 301: 62-69.

BINZ, H.R. and O. WILDI. 1988. *Das Simulationsmodell MaB-Davos.* Schlussberichte zum Schweizerischen MaB-Programm 33, Bern.

CLIFF, A.D. and J.K. ORD. 1981. *Spatial Processes: Models and Applications.* Pion, London.

DUTTER, R. 1985. *Geostatistik.* Teubner, Stuttgart. 159 p.

FEOLI E. and P. GANIS. 1986. Autocorrelation for measuring

predictivity in community ecology: an example with structural and chorological data from mixed forest types of NE Italy. Coenoses 1: 53-56.

FISCHER, H. 1990. Simulating distribution of plant communities in an alpine landscape. Coenoses 5: 35-41.

GREIG-SMITH, P. 1983. *Quantitative Plant Ecology*. 3rd. ed. Blackwell Scientific Publications, Oxford.

GOWER, J.C. 1971. Statistical methods of comparing different multivariate analyses of the same data. *In:* F.R. Hodson, D.G. Kendall and P. Tautu (eds.), *Mathematics in the Archeological and Historical Sciences*, pp. 138-149. Edinburgh University Press.

KELLER, M. and K. SEIDEL. 1984. Influence of Snow Cover Recession on an Alpine Ecological System. Proc. 18th Intern. Symposium on Remote Sensing of Environment (EIRM), Paris, France, Oct. 1-5: 1931-1936.

LEGENDRE, P. and M.-J. FORTIN. 1989. Spatial analysis and ecological modelling. Vegetatio 80: 107-138.

MANTEL, N. 1967. The detection of desease clustering and a generalized regression approach. Cancer Res. 27: 209-220.

ORLÓCI, L. 1987. *Multivariate Analysis in Vegetation Research*. 2nd ed. Junk, The Hague.

SILVERTOWN, J. 1980. The dynamics of a grassland ecosystem: Botanical equilibrium in the Park Grass Experiment. Journal of Applied Ecology 17: 491-504.

SOKAL, R.R. 1986. Spatial data analysis and historical processes. *In:* Diday, E. *et al.* (eds.), *Data analysis and Informatics, IV*. Proceedings of the Fourth International Symposium on Data Analysis and Informatics, Versailles, France, 1985, pp. 29-43. North-Holland, Amsterdam.

UPTON, G.J.G. and B. FINGLETON. 1985. *Spatial Data Analysis by Example*. Volume 1. *Point Pattern and Quantitative Data*. Wiley, Chichester.

WILDI, O. and K. EWALD (eds.). 1986. *Der Naturraum und dessen Nutzung im alpinen Tourismusgebiet von Davos*. Ergebnisse des MaB-Projektes Davos. Eidg. Anst. forstl. Versuchswes, Ber. 289.

WHITTAKER, R.H. 1978. Direct gradient analysis. *In:* R.H. Whittaker (ed.), *Ordination of Plant Communities*, pp. 7-50. Dr. W. Junk, the Hague.

Manuscript received: April 1989