



Quadrat size dependence, spatial autocorrelation and the classification of community data

J. Podani^{1,3} and P. Csontos²

¹ Department of Plant Taxonomy and Ecology, Eötvös University, Pázmány P. s. 1/c, H-1117 Budapest, Hungary

² Research Group of Theoretical Biology and Ecology, MTA-ELTE, Pázmány P. s. 1/c, H-1117 Budapest, Hungary

³ Corresponding author. E-mail: podani@ludens.elte.hu

Keywords: Clusterogram, Grasslands, Sampling, Variogram.

Abstract: The paper evaluates spatial autocorrelation structure in grassland vegetation at the community level. We address the following main issues: 1) How quadrat size affects the measurement of spatial autocorrelation for presence/absence and cover data? 2) What is the relationship between spatial autocorrelation and classification? 3) Is there a temporal change of spatial autocorrelation in the vegetation studied? We found that multivariate variogram shape, variance explained and the sill are different for presence/absence data and cover data, whereas quadrat size increases apparently introduce a stabilizing effect for both. Spatial stationarity is detected for species presence, and non-stationarity for cover. A new graphical tool, the clusterogram is introduced to examine spatial dependence of classification at various numbers of clusters. We found that spatial autocorrelation plays a crucial role in the classification of vegetation and therefore we suggest that its effect should not be removed from clustering. Mutual interpretation of variogram and clusterogram shape may be informative on the number of meaningful clusters present in the data. Spatial autocorrelation structure did not change markedly after 23 years for presence/absence data, indicating that the vegetation of the study area is stationary in time as well. The present study demonstrates that traditional quadrat data are suitable for evaluating spatial autocorrelation, even though field coordinates are recorded several years after sampling is completed.

Abbreviations: p/a – Presence-Absence, SA – Spatial Autocorrelation.

Introduction

Plant communities extend in a spatial continuum. As a consequence, the description and statistical analysis of compositional variation require quadrat data, unless precise maps are prepared on all plant positions. Quadrats (sampling units, plots) are delineated by the investigator in the two-dimensional space, with arbitrarily specified size, shape and orientation. The choice of these sampling parameters may strongly influence the results of numerical analyses and in turn the conclusions drawn by the ecologist (Kenkel et al. 1989). This problem has long been recognized in numerical ecology (see review in Podani 1984, and references therein), and the effect of quadrat size, in particular, exemplifies the generally recognized issue of spatial scaling (Wiens 1989, Jelinski and Wu 1996). Further complication is that statistical analyses supplemented with significance tests are burdened in most cases by the lack of independence among quadrats

(Dale and Fortin 2002, Legendre et al. 2002). The fact that data pertaining to a quadrat are correlated with those obtained from quadrats taken in the physical neighborhood, and that this relationship depends on between-quadrat field distances are manifestations of a phenomenon widely known as “spatial autocorrelation” (abbreviated as SA; Legendre and Fortin 1989, Legendre 1993, Legendre and Legendre 1998, Koenig 1999, Perry et al. 2002, Wagner 2003).

Bibliographic surveys suggest that SA has been in the research focus of several branches of ecology. In the overwhelming majority of studies, emphasis is placed on univariate situations in which autocorrelation of a single variable, such as the abundance of a given species, is examined in the function of between-quadrat distances. Moran's *I* statistic and Geary's *c* function are just two examples from the extensive methodological arsenal introduced for this purpose (Legendre and Legendre 1998,

Perry et al. 2002). Community data are basically of multivariate nature, however, to which this standard methodology is unsuitable, so that special approaches are called for. Suggested modifications to community data include the calculation of some (dis)similarity coefficient between all pairs quadrats or sites (Jaccard index: Nekola and White 1999, Mistral et al. 2000; squared Euclidean distance: Wagner 2003) and the display of (dis)similarities against spatial inter-quadrat distances. Instead of showing (dis)similarities directly, the Mantel statistic is derived from the comparison of the dissimilarity matrix and a descriptor matrix representing a given distance class, and then these statistics are depicted in the so-called Multivariate Mantel Correlogram (Oden and Sokal 1986, Legendre and Fortin 1989, Legendre and Legendre 1998). Another indirect procedure involves calculating spatial autocorrelation from ordination scores derived from community data (Anand and Kadmon 2000). Unfortunately, while the paradigm shift towards spatial autocorrelation opened up new perspectives in ecological data analysis, the classical problems associated with quadrat sampling are almost completely neglected. The methodological contrast becomes clear if we consider that there have been only three reports on the effect of quadrat size and shape upon the estimation of spatial autocorrelation in plant communities (Bellehumeur et al. 1997, Fortin 1999, Maestre et al. 2005). However, these studies are restricted to pooled abundances irrespective of species identity or to separate analyses of the species or species pairs and are therefore *univariate* or *bivariate* in nature. The effect of quadrat size upon estimating spatial autocorrelation at a truly *community level* has remained unexplored. One objective of this paper is to examine this problem in grassland vegetation, based on both presence/absence and cover data.

Whereas it is generally accepted that SA cannot be ignored in significance tests (Fortin and Jacquez 2000), the relationship of SA to community classification is largely unclear. If one wishes to consider spatial proximity of quadrats in classifying them, then constrained clustering offers various solutions (Legendre 1987, Gordon 1999). To the authors' knowledge, constraints based on geostatistical functions were only applied by Bourgault et al. (1992). In their approach, large distances between geological sites that were closely located were down-weighted, thus influencing the results of cluster analysis. In the classification of plant communities, the role of SA and the question whether SA should be incorporated at all remain to be investigated, even though Legendre (1987) put forward the suggestion that autocorrelated data should always be classified via constrained clustering. Using the grassland data, we explore the spatial dependence of

vegetation classification and examine this theoretical issue in some detail.

The third question that deserves attention is related to the applicability of quadrat data, obtained through less regular sampling designs via classical phytocoenological methods in apparently inhomogeneous vegetation stands, to evaluating SA. Most published results on SA structures rely upon systematic arrangements of quadrats, such as grids (Legendre and Fortin 1989, Fortin 1999, Wagner 2003, Maestre et al. 2005) or line transects (Mistral et al. 2000) as well as upon sites separated by large biogeographical distances (Nekola and White 1999). In this study, we utilize data from quadrats arranged in an irregular manner within a relatively small study area, such that accurate geographical positions were determined several years after field work was done (explained in the Methods section). In other words, sampling design was not chosen deliberately to provide data for SA analysis. This issue is important because very large amounts of data have accumulated in vegetation ecology based on irregular and subjective designs, whose *a posteriori* evaluation with respect to SA may extend our ecological knowledge considerably.

Plant communities are subject to change over time and it is an interesting question if these changes are associated with SA changes as well. Succession implies replacement of quite distinct vegetation types by one another with definite directionality. On the other hand, communities that appear in temporal equilibrium are also fluctuating without any specific direction of change. We may assume with good reason that successional trends strongly modify the autocorrelation structure, for example, in a transitional sere from bare land to forest. We cannot predict, however, if the temporal component of short-term spatiotemporal community changes is large enough to influence SA. We explore this problem for the presence/absence case, for one quadrat size and for two points of time 23 years apart, in the grassland site to provide the first example of evaluating temporal SA changes in a vegetation type with fluctuating species composition.

Material

Study area

Field work was made in the rock grassland vegetation of the Sas-hegy Nature Reserve situated within the city limits of Budapest, Hungary. The reserve is of 30 ha area separated by fence from the surrounding built-in urban land. Sas-hegy is composed of three main outcrops, the highest reaching an elevation of 259 m above sea level. The slopes, especially the south facing ones are steep,

with an inclination of 45° or more. The bedrock is Triassic dolomite, a soft rock subject to intensive superficial erosion of mostly physical nature.

Vegetation

The vegetation of Sas-hegy is a mosaic of natural, semi-natural and secondary vegetation fragments. The present study is confined to the less disturbed natural stands of rock grasslands which cover approximately one-third of the total area of the entire reserve. Thus, the sampled area has the approximate dimensions of 400 m by 300 m. The dominant species of these communities include *Festuca pallens* Host, *Carex humilis* Leyss., *Chrysopogon gryllus* (Torn.) Trin., and *Sesleria sadleriana* Janka. Extensive ordination and classification studies based on presence/absence data revealed that the vegetation of Sas-hegy represents a spatial continuum, although several quadrats have the tendency of being classified together in so-called *nodal groups* (*sensu* Noy-Meir 1974), irrespective of sample plot size, dissimilarity function and clustering strategy used (Podani 1998). The observation that the Sas-hegy grasslands exhibit a group structure as well as a continuum simultaneously makes this object extremely suitable for comparing SA analyses and classifications.

Methods

Sampling

A detailed sampling survey was made by the first author from March to September, 1977. A total of 80 quadrats were placed such that areas obviously degraded or overgrown by naturalized ornamental shrubs (mostly *Syringa vulgaris* L.) were discarded. Fairly equal representation of the less disturbed parts of the vegetation was ensured by placing the quadrats evenly yet irregularly in the field. Each square plot was a nested system of 8 small subquadrats increasing in size (Table 1), with a common corner. Percentage cover of vascular plant species was estimated by eye for each size. These data will provide the basis for evaluating spatial dependence of dissimilarity and classification.

Sampling was repeated in 2000 (Podani et al. 2005). The quadrats were reestablished using field marks and mapped positions (semipermanent plots, *sensu* Persson 1980) and were visited several times from April to September, to record presence/absence scores in 4 × 4 m² quadrats. This simplified sampling design was chosen according to the results of a complex multivariate experiment analyzing the relative effect of plot size and data type on classifications and ordinations (Podani 1989).

The choice in favor of presence/absence data was supported by lower sampling effort, decreased sampling error and the observation that plot size effects were considerably smaller in case of presence/absence data than when cover scores were used. Podani et al. (2005) selected the largest quadrat size because the analyses indicated a trend towards stabilization of results along with quadrat size increases.

Determining geographic positions

The quadrat positions were identified in the field using notes and maps made in 1977. In the centre of each 4 × 4 m² quadrat, the geographical position was determined using global positioning system (GPS) with a Garmin-12 receiver. Average time spent in acquiring data was 3–4 minutes at each quadrat and it has a point accuracy of 2 m, which is not larger than the distance of the 4 × 4 m² quadrat centre from the centre of the 0.25 × 0.25 m² quadrats. Therefore, we used the same coordinates for all quadrat sizes in all analyses. Elevation was also determined by the same GPS technology. The two-dimensional spatial coordinates plus the elevation data were expressed in meters and collected in an 80 × 3 matrix **P**, allowing the comparison of results based on two and three dimensions.

Data analysis

Multivariate spatial autocorrelation was evaluated according to standard methodology as described in Wagner (2003). The essence of the SA analysis is that half the squared difference between the vegetation of two quadrats (the semivariance) is plotted against their distance in space. In order to avoid confusion, “distances” will refer in this paper only to between-quadrat dissimilarities calculated from vegetation data and “lag distances” to the separation distances of quadrats in the physical space. Since this would lead to a scatterplot that is usually difficult to evaluate, the semivariances are averaged for a series of lag distance classes to yield the so-called empirical multivariate variogram. That is, the semivariance for lag distance class *h* is obtained from the squared Euclidean distances of all pairs of quadrats *j* and *k* falling into that category, namely

$$S(h) = \frac{1}{2} n_h \sum_i \sum_{j < k | d_{jk} \in A(h)} (x_{ij} - x_{ik})^2 \quad (1)$$

where *A(h)* is the set of lag distances *d_{jk}* in category *h* with *n_h* elements, and *x_{ij}* is the performance (presence/absence, cover, etc.) of species *i* in quadrat *j*. The variogram is called omnidirectional, because there is no attention paid to the direction of lag distances. For presence/absence data, *S(h)* is the mean number of species that are present

only in one of a pair of quadrats in $A(h)$ and is called the *complementarity* of species composition of two quadrats (Wagner 2003). In other words, $S(h)$ is half the mean Manhattan distances between quadrats within lag distance category h . The *variogram* reflects the SA structure of the community and is characterized by three important parameters. The *range* is the lag distance where the variogram becomes constant. Beyond the range, two quadrats can be considered to be independent observations whereas within the range the closer are two quadrats, the more similar their vegetation. The value of $S(h)$ for lags beyond the range is the *sill*, corresponding to the half of mean squared distances between independent quadrats. The third parameter is the *nugget*, the estimated value of $S(h)$ for zero lag distances, which is the unexplained variance at a scale below the resolution of the analysis, including within quadrat variation and measurement error. The relative importance of the nugget variance is obtained by dividing it with the sill to calculate the proportion of variance unexplained by the variogram. This ratio has been termed the *relative nugget effect* (e.g., Solie et al. 1996). If it is near 1, then a high proportion of variability is unexplained, and therefore there is a weak relationship between quadrats even if they are closely spaced. When the relative nugget effect approximates zero, then quadrats within the range are strongly correlated regarding the variables measured.

Variograms were computed for all quadrat sizes separately based on cover and presence/absence scores obtained in 1977 to evaluate the effects of sampling scale. For assessing temporal changes in the SA structure, a variogram for the p/a data from $4 \times 4 \text{ m}^2$ quadrats surveyed in 2000 was also calculated to allow comparison with the corresponding variogram for 1977. Lag distances computed from **P** were assigned into 18 categories, 20 m wide each. The variogram data were calculated by FORTRAN programs written by the first author.

The relationship between classification and SA was evaluated in a complex manner to explore how clustering tendency is influenced by geographical distances among quadrats. The clustering strategy for classifying quadrats based on both data types was the incremental sum of squares (Ward) method (see Podani 2000) which is statistically compatible with the sum of squares-based SA analysis. We selected the largest quadrat size to allow comparison of two study years in the presence/absence case, and we used the same size for cover data from 1977. Cluster analyses were performed by the SYN-TAX 2000 package (Podani 2001). The number of clusters examined was 2-6 in all cases, the partitions were obtained by cutting horizontally the dendrograms at the 1st to 5th hierar-

chical level from the top, respectively. According to previous studies, this range of partitions includes the ecologically most meaningful classifications of the studied vegetation (Podani 1998) and therefore no further refinement was made into more clusters. The relationship between classification and lag distances was examined graphically. For each category, we calculated the ratio of the number of within-cluster distances to the total number of distance values falling into that category. This ratio is plotted as a function of h to yield a graph, to which we coin a new name, the *clusterogram*. That is, the clusterogram is defined as

$$R(h) = |B(h)| / |A(h)| \quad (2)$$

where $A(h)$ is defined as above, $B(h) \subseteq A(h)$ is the set containing within-group distances in the classification at a specified number of clusters, and $||$ indicates the cardinality of (= the number of elements in) the sets. $R(h)$ is thus a cardinality ratio reflecting *clustering tendency* of quadrats within a class of lag distances. The clusterograms are displayed for the same set of lag distance categories as the semivariograms.

Results and discussion

There was no noticeable difference between the variograms based on the two-dimensional spatial coordinates only and on spatial positions with elevation included as well. Partitioning the total variance by dimensions revealed that elevation accounted for only 1.7% of variance in the distances, so that the entire discussion below is confined to the two-dimensional case. The number of quadrat pairs per category of lag distances ranged from 2 to 293, with 93 values in the first category. We decided to omit those large lag distance categories in which the number of values was smaller than 93, to diminish distortion of averages for large lag distances caused by spatial constraints and border size effects (Wagner 2003). In the last four lag distance categories, the values were 76, 30, 15 and 2, so these classes were omitted from future analyses. In other words, lag distances larger than 280 m were discarded. In the remaining 14 categories, the numbers of values run from 93 to 273, providing a reliable basis for calculating variograms and clusterograms. The mean value within the first lag distance category was 13.4 m, so this is on the average the lowest resolution achieved in the present study.

Quadrat size dependence of SA

The spatial autocorrelation structure differed with quadrat size. In the presence/absence case (Fig. 1.a), there

was a slight, almost monotone increase of the semivariance at the smallest quadrat size, which is usually interpreted as a result of spatial trend (drift or gradient) in geostatistics (Royle et al. 1980). This unbounded variogram indicated that variance was not constant for fine grained plant pattern (at a scale of 0.25 m^2). For $1 \times 1 \text{ m}^2$ and more, however, the variograms were bounded, indicating a constant variance (i.e., half squared pairwise distance) over the study region. The range of the variogram appeared around lag distance category 5, which corresponds to 80–100 m. Quadrats separated by larger lag distances can therefore be considered independent in species composition. Variogram shape apparently converged to a final one, as shown by the very similar curves obtained for the

five largest quadrat sizes. The sill, i.e., the number of species pertaining on the average to one of the quadrats compared, increased along with quadrat size, from ca. 8 to 16. Table 1 summarizes sill values estimated using the formula in which p_i is the estimated probability of presence for species i in a random quadrat (Wagner 2003). Sill increases were proportional to increases in the average number of species per quadrat (Table 1). For the largest quadrats, sill was half the average Manhattan distance between quadrats (31.98) reported already from a previous study (Podani et al. 2005). The interval of the semivariations was narrowed towards small lag distance classes, so that the nugget values ranged only from ca. 7 to 10. The relative nugget effect was maximum for the smallest

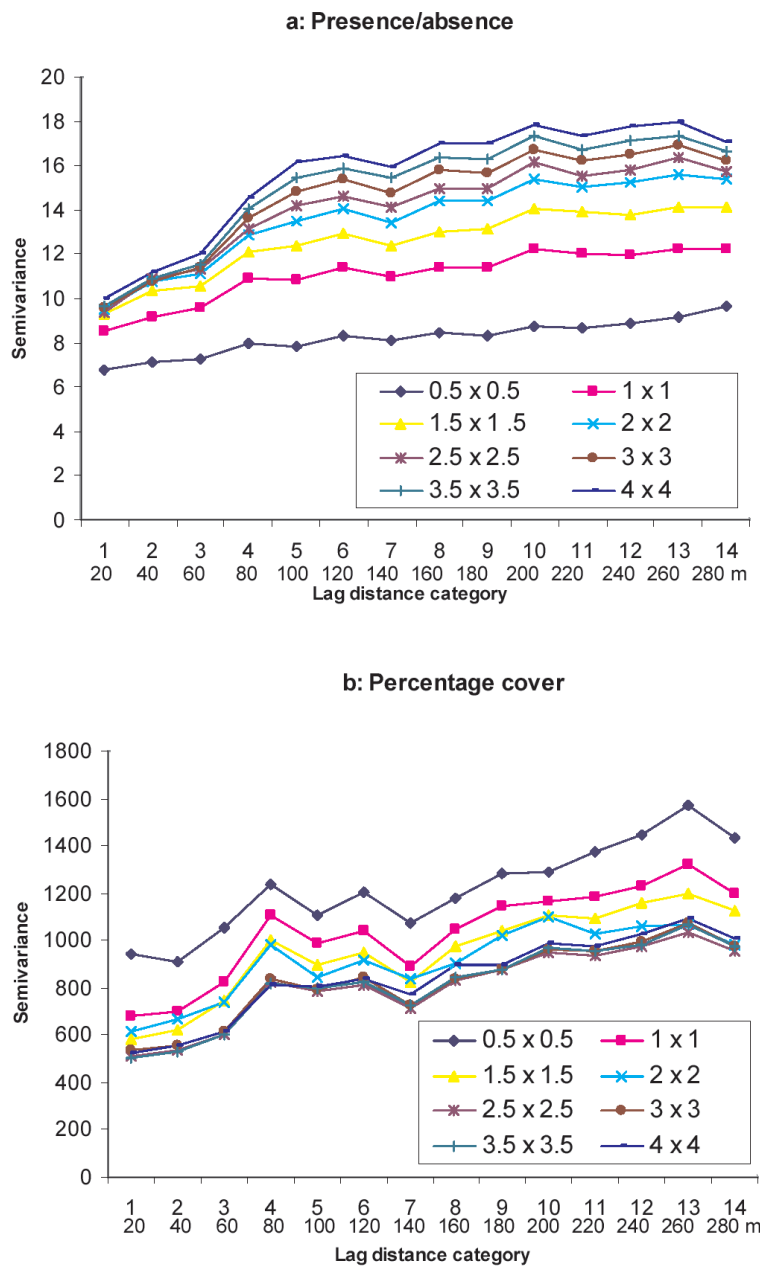


Figure 1. The effect of quadrat size on the variograms for two data types based on the 1977 data. Quadrat size is given in $\text{m} \times \text{m}$ in the legend box.

Table 1. Dependence of species richness and estimated sill on quadrat size in the Sashegy grassland vegetation in 1977.

Quadrat size (m ²)	Average number of species per quadrat	Total number of species found	Estimated value of sill for presence/absence
0.5 x 0.5	12.5	79	8.2
1 x 1	19.4	101	11.1
1.5 x 1.5	24.7	108	12.6
2 x 2	28.6	110	13.7
2.5 x 2.5	31.6	112	14.2
3 x 3	34.8	118	14.8
3.5 x 3.5	37.4	121	15.2
4 x 4	39.8	123	15.8

quadrat size (0.85) indicating that high proportion of the variance remained unexplained if the grain of the survey was fine. This ratio, however, gradually decreased over quadrat size, reaching 0.6 for $4 \times 4 \text{ m}^2$.

The variograms had very different appearances and therefore different interpretations for cover data (Fig. 1.b). The shape of the variograms and the interval of semi-variances were much less influenced by quadrat size. In contrast with the presence/absence case, the semivariance was the highest for the smallest quadrat size. A remarkable similarity between variogram sets obtained from the two data types was the convergence of curves to a “mean shape”. For the largest four quadrat sizes, the variograms were almost indistinguishable from one another. This had to do with the common observation that very small quadrats give highly variable estimates of percentage cover, and quadrat size increases introduce a stabilizing effect. The contrasting tendency that the number of species increased along with quadrat size was thus overwhelmed by decreases in the random fluctuations of cover. The overall SA structure, however, was less clear-cut than in the p/a case. The variograms appeared to level off after reaching the 4th lag distance category, and to fluctuate around a “sill” for the next four categories, but then the semivariance increased again suggesting a trend for large lags. The lag distance where the upward trend begins, the *departure distance* (Solie et al. 1996) is not calculated usually in semivariogram analysis, but deserves attention in the present case. This phenomenon may indicate the existence of a nested spatial structure or lack of stationarity: variance was constant for quadrats up to 160 m apart, but beyond this we encountered further increases in the variance. That is, based on percentage cover the entire vegetation could be subdivided into strata with different expectations for between-quadrat distances within each type. Each of these strata could be characterized by the dominance of a

few species and it was possible that site conditions also differed with strata. This is in sharp contrast with the presence/absence case for which we can assume second-order stationarity over all the study area. An alternative explanation of departure distance is the border effect which may influence the results when the range (in this case 160 m) is relatively large compared to field size (ca. 300 m by 400 m). Since we have no possibility to expand the study region any further, the reasons behind background causes of departure distance cannot be revealed with certainty.

Analogously to the p/a case, the semivariance may be interpreted as half the squared cover values in which two quadrats differ. We found that for the largest quadrat size a total cover of 30% pertained to each quadrat on the average. The sill decreased from 1300 to 900, and the nugget from 900 to 500 along with quadrat size increments. The relative nugget effect therefore decreased from 0.67 to 0.55, a tendency also observed in the presence/absence situation although in this case the change of nugget effect was smaller. For the largest quadrats, the relative nugget effect was approximately the same for the two data types. That is, around 55-50% of the variation remained unexplained by the present analysis. Part of the unexplained variation came from within-quadrat fluctuations (i.e., noise), while the other part from sampling error. The very high relative nugget effect for the smallest quadrats, and field experience suggest that within-quadrat fluctuations dominated.

Temporal change of SA

The variograms for the presence/absence data and quadrat size of $4 \times 4 \text{ m}^2$ are shown in Figure 2 for the two study years. The curves are essentially of the same shape, with the exception that the 2000 curve is above the 1977 curve for all lags. It means that the sill and the nugget increased slightly whereas the range remained approxi-

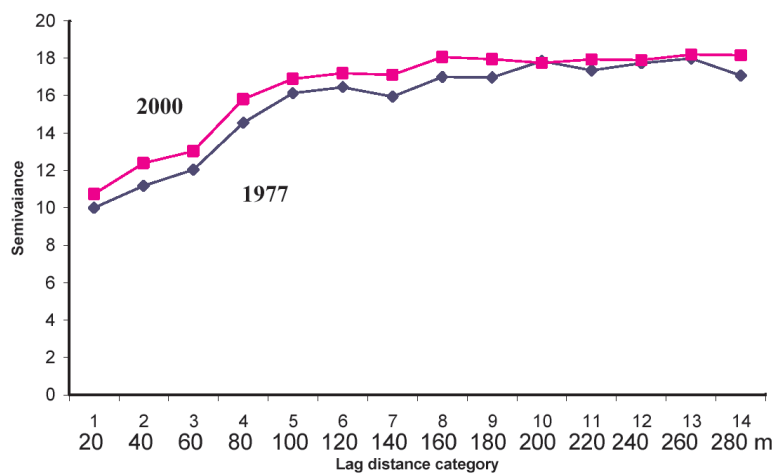


Figure 2. The temporal changes of spatial autocorrelation structure for presence/absence data and $4 \times 4 \text{ m}^2$ quadrat size in the two study years.

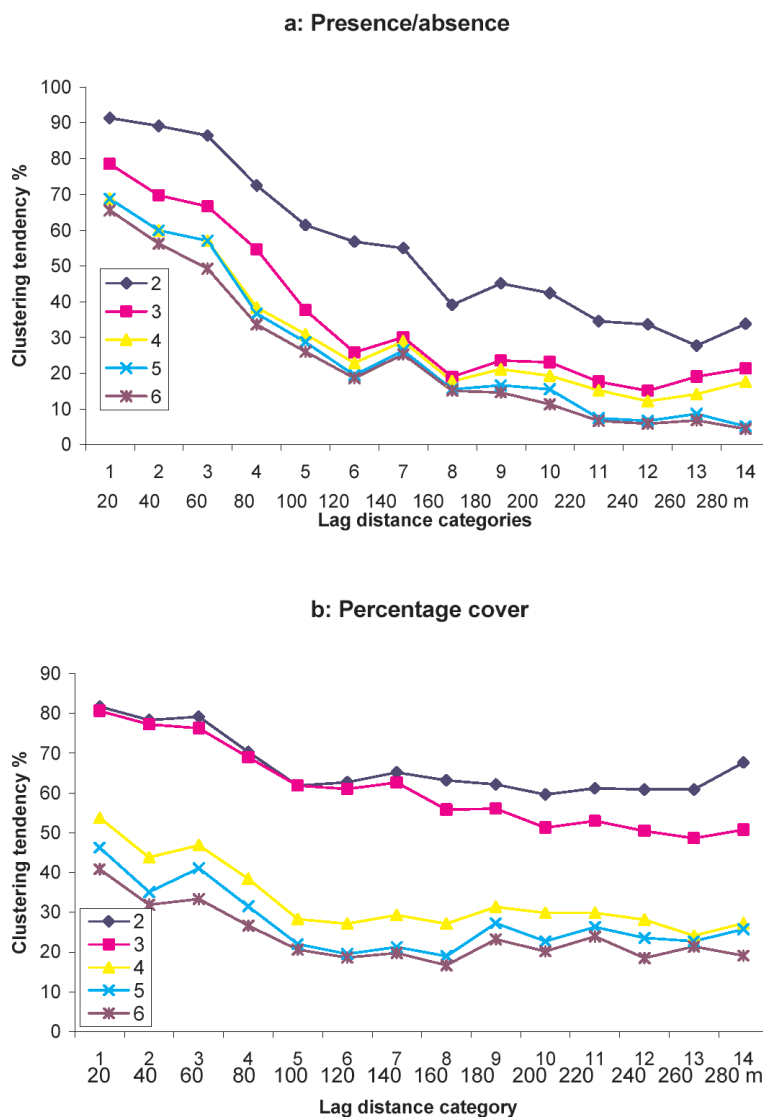


Figure 3. Clusterograms showing clustering tendency of $4 \times 4 \text{ m}^2$ quadrats for two data types and 2-6 clusters (set insets, for explanation of symbols) across lag distances 0-280 m based on the 1977 data.

Table 2. List of species with the highest discriminatory power at the three-cluster classification of cover data. Each value corresponds to the mean cover percentage of the species in the given group of quadrats.

Species	Cluster A (3 quadrats)	Cluster B (17 quadrats)	Cluster C (60 quadrats)
<i>Sesleria sadleriana</i> Janka	60.0	2.6	0.2
<i>Scabiosa canescens</i> W. et K.	9.3	7.5	1.4
<i>Anthericum ramosum</i> L.	7.3	3.1	0.8
<i>Bromus erectus</i> Huds.	9.6	39.8	8.3
<i>Carex humilis</i> Leyss.	4.3	6.6	0.7
<i>Festuca pallens</i> Host	3.6	3.5	11.7
<i>Fumana procumbens</i> (Dun.) Gren. et Godr.	1.0	4.9	10.1
<i>Seseli leucospermum</i> W. et K.	0.6	2.2	6.2
<i>Chrysopogon gryllus</i> (Torn.) Trin.	0	0.1	4.3
<i>Carex liparicarpus</i> Gaud.	0	0	1.7

mately the same after 23 years, the increases explained by the higher species richness in 2000 (139 species versus 123 species in 1977). This observation corresponds well with a comparative study of the same vegetation based on the 1977 and 2000 data at the level of data and of classifications and ordinations (Podani et al. 2005): whereas increases of species richness were detected along with considerable rearrangements in the results, the overall spatial structure of the grassland vegetation remained fairly constant.

Spatial dependence of classification

The clusterograms for presence/absence data from the first study year (Fig. 3.a) indicated that clustering tendency had a strong spatial dependence; it decreased with lag distance almost monotonically. For two clusters (top curve in the diagram), around 90% of the distances falling into lag distance categories 1-3 were within-cluster distances in the classification, whereas this percentage dropped to around 30 for the last three classes. This was a straightforward indication that spatial closeness played crucial role in the classification of vegetation. This may be interpreted as a relatively weak tendency to exhibit repeated presence/absence pattern in remote distances within the study area. The fact that the percentages decreased monotonically for all lag distances over increases in the number of clusters is a mathematical necessity, since breaking clusters up produces more and more between cluster distances. The large gap between the top curve and the others, which are fairly close to one another, has also to do with the properties of the underlying dendrogram. The gap was caused by splitting a large cluster into approximately equal halves. Further changes in the

shape and position of curves were less dramatic, because the further splits produced clusters strongly unequal in size.

The clusterograms for cover data (Fig. 3.b) differed from those in Fig. 3.a in shape and arrangement as well. Most curves were similar in shape, with an initial decreasing phase in clustering tendency, which seemed to stabilize for large lag distances. For 2 and 3 clusters, about 80% of distances pertaining to small lag distances were within-cluster distances, a slightly smaller figure than for p/a data. However, clustering tendency remained around 50-60% for all subsequent categories. Similar statements hold true for the other 3 curves: clustering tendency started at 42-53%, and remained around 20-30% for lag distance categories 5-14. This suggests that clustering of quadrats based on cover data was less dependent on spatial separation than in the presence/absence case. The exception is classification into 3 clusters: in the associated clusterogram we found a definite decrease of clustering tendency throughout the lag distances. The gap between the two groups of curves was indicative of splitting a large cluster into two groups of similar size.

The variogram for quadrat size $4 \times 4 \text{ m}^2$ in Fig. 1.b and the clusterogram for 3 clusters in Fig. 3.b deserve particular attention. After the departure distance in the variogram (lag 8) the corresponding clusterogram was still in a decreasing phase, which was not so with other numbers of clusters. In other words, for the scale in which non-stationarity holds, clustering tendency was still decreasing. We can consider this phenomenon as an indication of the optimum number of clusters in the classification. This was so even though there was a large group in the dendro-

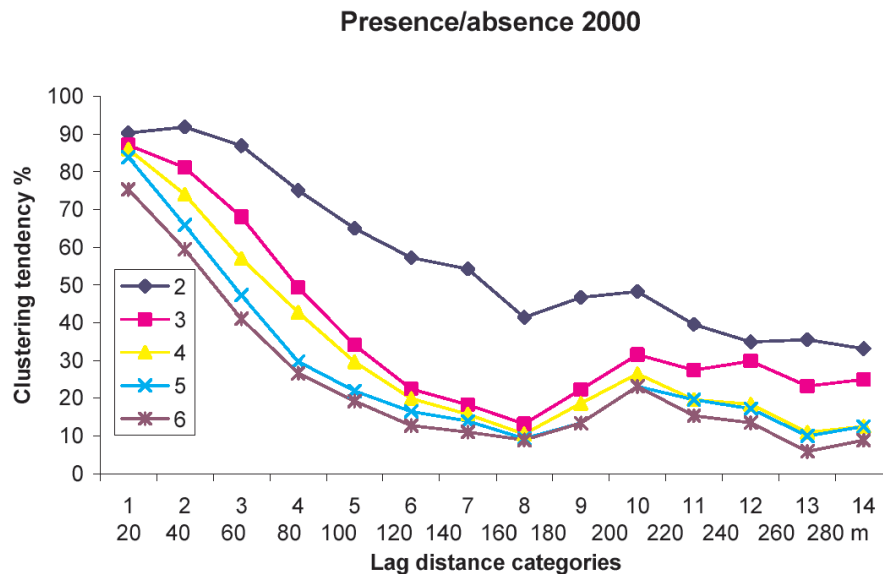


Figure 4. Clusterograms showing clustering tendency of $4 \times 4 \text{ m}^2$ quadrats for presence/absence data and 2-6 clusters (set inset, for explanation of symbols) across lag distances 0-280 m based on the 2000 data.

gram with 60 quadrats, whereas the other two clusters contained only 3 and 17. Since the analysis was based on Euclidean distances from percentage cover without standardization, it was easy to identify the species that dominate separation (Table 2). We note that most of the species, which played crucial role in the presence/absence case, were not influential now.

Comparison of clusterograms for presence/absence data in the two study years (Figures 3.a and 4) showed that spatial dependence of clustering did not change substantially after 23 years. The corresponding curves were similar in shape, and the gap between the 2- and 3-cluster cases remained wide. This was so even though the classification itself had changed remarkably, at the three-cluster level, for example, 17 objects were relocated from one cluster to the other. Again, this was in good agreement with our previous conclusion (Podani et al. 2005) that rearrangements in the classification were not associated with changes in the basic spatial characteristics of the vegetation studied.

Closing remarks and future perspectives

The present paper has shown for the first time that our perception of multivariate spatial autocorrelation in vegetation depends on quadrat size for both percentage cover and presence/absence data. The nugget/sill ratio decreases along with quadrat size increases, showing that the spatially structured component of variance becomes larger. Whereas these observations may appear of general relevance because similar findings were reported from univariate studies as well (Bellehumeur et al. 1997), some

details are study-specific. The difference between the variograms pertaining to different data types is a fundamental property of the vegetation studied. In the present study, the bounded variograms obtained for the presence/absence data at large quadrat size indicate that the entire vegetation is statistically the same “population” with a common variance (i.e., pairwise Manhattan distance). In contrast, the unbounded variograms derived for the cover data reflect unequal variances over the entire region, suggesting either the presence of embedded large community patches within which the variance can be considered constant, or increased border effects. Stationarity of vegetation is therefore data-dependent. The similarity of the relative nugget effect for the two data types obtained at the largest quadrat size is probably also case-dependent. The convergence of shape to an “expectation” along with increases of quadrat size implies that the range of autocorrelation is stable except for small quadrats, while univariate studies appear less conclusive in this regard. For example, the range diminished (Bellehumeur et al. 1997) or increased (Maestre et al. 2005) for the density of all trees or chamaephytes, respectively, whereas there were no noticeable changes in correlogram shape for species abundances evaluated separately (Fortin 1999).

We introduced a new graphical tool, the clusterogram, for the assessment of the spatial dependence of clustering tendency among quadrats. The shape of the curves is strongly influenced by data type, and the properties of the associated dendrogram are also influential. In any case, most of the within-cluster distances fall into small lag distance categories, showing that the classification is greatly influenced, or more precisely, it is largely determined by

spatial autocorrelation. We feel, therefore, that compensating this effect by some sort of constrained clustering would produce a less interpretable or even meaningless classification of quadrats. Whereas spatial autocorrelation requires special treatment whenever significance tests are to be performed on the data, which assume statistical independence of the observations, clustering imposes no restrictions in this regard. In general, procedures with the objective of pattern exploration (cf. Kenkel et al. 1989) do not assume independence. There is no theoretical justification for not classifying spatially dependent quadrats together, because *it is* spatial dependence, i.e., a common generating mechanism in the background, that puts them into the same cluster. The situation is somewhat analogous to phylogenetic classification: in that case common genealogy (phylogenetic autocorrelation) corresponds to spatial autocorrelation, taxa replace quadrats and the temporal continuum of organisms corresponds to the topographic space.

Mutual interpretation and comparison of semi-variograms and clusterograms allow further insight into the spatial properties of the vegetation studied. Quadrats in the same cluster are on the average more autocorrelated in the presence/absence case than for cover data. Clustering tendency beyond the range is very weak for presence/absence, i.e., the spatial component in classification is strong. For cover data, however, the spatial component appears much less influential with respect to clustering: after reaching the departure distance clustering tendency remains the same for very large lags with the exception of 3-cluster classification. Further studies in different vegetation types and using simulated data are required to confirm or reject the general validity of these observations and to assess the background processes behind these phenomena.

If spatial autocorrelation is considered an unavoidable component of vegetation classification, some new analytical possibilities arise. For example, Shing (2003) raises the possibility of detecting the optimum number of clusters in a classification by examining spatial autocorrelation. The classification results are superimposed on a map and clusters that are larger in diameter than the range are split into two, a process continued until all diameters become shorter than the range. The question whether this approach, suggested originally to landscape level studies, is suitable to areas with smaller extent is another possibility that deserves attention in the future. Our study was not suitable to this kind of analysis, because the sampling units were located in vegetation fragments that were intermingled with patches greatly affected by disturbance. We have a different proposition, however, that arises from

comparisons between variograms and clusterograms. We suggest that for the optimum number of clusters, variogram and clusterogram shapes must coincide in some sense and nevertheless feel that further studies are required to confirm this view. In our case, the trend appearing after the departure distance and the decrease of clustering tendency for the same lags were considered as indicative of the “correct” number of clusters.

The temporal changes of variograms and clusterograms suggest that the grassland vegetation of Sas-hegy is in spatial stationarity. At least in the presence/absence case, the structure of the communities is unchanged in time, confirming results from an extensive classification and ordination study (Podani et al. 2005).

Much experience has accumulated on the basic properties of semivariograms, on their interpretation and on the causes behind their departures from common models (see Legendre and Legendre 1998, and references therein). This is not so, of course, with the newly introduced clusterogram procedure. The present investigation revealed that most properties of clusterograms reflect spatial aspects whereas others tend to reflect the underlying classification, and that the interplay of these two may give various curves with different dependence on the number of clusters. For example, there is correspondence between clusterogram shape and the manner clusters are split into smaller groups in the dendrogram. The effects of spatial properties, e.g., patchiness, periodicity and trend, on clusterograms require further studies on actual and artificial data sets.

Acknowledgements. Financial support from Hungarian Research Grants OTKA T43732 for J. Podani and OTKA T037732 for P. Csontos is greatly acknowledged. We are grateful to the anonymous referees for their comments on the previous versions of this manuscript. Thanks are due to the management of the Danube-Ipoly National Park for granting access to the study area in Sas-hegy.

References

- Anand, M. and Kadmon, R. 2000. Community-level analyses of spatiotemporal plant dynamics. *Écoscience* 7: 101-110.
- Bellehumeur, C., Legendre, P. and Marcotte, D. 1997. Variance and spatial scales in a tropical rain forest: Changing the size of sampling units. *Plant Ecol.* 130: 89-98.
- Bourgault, G., Marcotte, D. and Legendre, P. 1992. The multivariate (co)variogram as a spatial weighting function in classification methods. *Math. Geol.* 24: 463-478.
- Dale, M. R. T. and Fortin, M.-J. 2002. Spatial autocorrelation and statistical test in ecology. *Écoscience* 9: 162-167.
- Fortin, M.-J. 1999. Effects of sampling unit resolution on the estimation of the spatial autocorrelation. *Écoscience* 6: 636-641.
- Fortin, M.-J. and Jaquez, G. M. 2000. Randomization tests and spatially autocorrelated data. *Bulletin of the ESA* 81: 201-206.

- Gordon, A. D. 1999. *Classification*. 2nd ed. Chapman and Hall, London.
- Jelinski, D. E. and Wu, J. 1996. The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecol.* 11: 129-140.
- Kenkel, N. C., Juhász-Nagy, P. and Podani, J. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195-207.
- Koenig, W. D. 1999. Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.* 14: 22-26.
- Legendre, P. 1987. Constrained clustering. In: Legendre, P. and Legendre, L. (eds), *Developments in Numerical Ecology*. Springer, Berlin, pp. 289-307.
- Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology* 74: 1659-1673.
- Legendre, P., Dale, M. R. T., Fortin, M.-J., Gurevitch, J., Hohn, M. and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25: 601-616.
- Legendre, P. and Fortin, M.-J. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80: 107-138.
- Legendre, P. and Legendre, L. 1998. *Numerical Ecology*. 2nd ed. Elsevier, Amsterdam.
- Maestre, F. T., Rodríguez, F., Bautista, S., Cortina, J. and Bellot, J. 2005. Spatial associations and patterns of perennial vegetation in a semi-arid steppe: a multivariate geostatistics approach. *Plant Ecol.* 179: 133-147.
- Mistral, M., Buck, O., Meier-Behrmann, D. C., Burnett, D. A., Barnfield, T. E., Scott, A. J., Anderson, B. J. and Wilson, J. B. 2000. Direct measurement of spatial autocorrelation at the community level in four plant communities. *J. Veg. Sci.* 11: 911-916.
- Nekola, J. C. and White, P. S. 1999. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26: 867-878.
- Noy-Meir, I. 1977. Multivariate analysis of the semiarid vegetation in south-eastern Australia. II. Vegetation catenae and environmental gradients. *Australian J. Bot.* 22: 115-140.
- Oden, N. L. and Sokal, R. R. 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Syst. Zool.* 35: 608-617.
- Persson, S. 1980. Succession in a south Swedish deciduous wood: a numerical approach. *Vegetatio* 43: 103-122.
- Perry, J.N., Liebhold, A. M., Rosenberg, M. S., Dungan, J., Miriti, M., Jakomulska, A. and Citron-Pousty, S. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography* 25: 578-600.
- Podani, J. 1984. Spatial processes in the analysis of vegetation: theory and review. *Acta Bot. Hung.* 30: 75-118.
- Podani, J. 1989. Comparison of classifications and ordinations of vegetation data. *Vegetatio* 83: 111-128.
- Podani, J. 1998. A complex numerical analysis of dolomite rock grasslands of the Sas-hegy Nature Reserve, Budapest, Hungary. In: Csontos, P. (ed.), *Botanical Exploration of Rock Grasslands. A Festschrift to B. Zólyomi*. Scientia, Budapest, pp. 213-229 (in Hungarian, with English abstract).
- Podani, J. 2000. *Introduction to the Exploration of Multivariate Biological Data*. Backhuys, Leiden.
- Podani, J. 2001. *SYN-TAX 2000. User's Guide*. Scientia, Budapest.
- Podani, J., Csontos, P., Tamás, J. and Miklós, I. 2005. A new multivariate approach to studying temporal changes of vegetation. *Plant Ecol.* 181: 85-100.
- Royle, A., Clark, I., Brooker, P. I., Parker, H., Journel, A., Rendu, J. M., Sandefur, R. I. and Mousset-Jones, P. 1980. *Geostatistics*. McGraw-Hill, New York, N. Y.
- Shing, Lin. 2003. Optimization of cluster analysis using autocorrelation. Department of Geography, Southwest Texas State University. Presented at UCGIS Summer Assembly, Pacific Grove, CA, 2003. Manuscript available from www.ucgis.org/summer03/studentpapers/shinglin.pdf.
- Solie, J. B., Raun, W. R., Whitney, R. W., Stone, M. L. and Ringer, J. D. 1996. Optical sensor based field element size and sensing strategy for nitrogen application. *Transactions of the American Society of Agricultural Engineers* 39: 1983-1992.
- Wagner, H. H. 2003. Spatial covariance in plant communities: Integrating ordination, geostatistics, and variance testing. *Ecology* 84: 1045-1057.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Funct. Ecol.* 3: 385-397.

Received May 2, 2006
Accepted May 17, 2006