Folia Geobotanica 42: 141–152, 2007

STATISTICAL AND BIOLOGICAL CONSEQUENCES OF PREFERENTIAL SAMPLING IN PHYTOSOCIOLOGY: THEORETICAL CONSIDERATIONS AND A CASE STUDY

Zoltán Botta-Dukát*), Edit Kovács-Láng, Tamás Rédei, Miklós Kertész & János Garadnai

*Institute of Ecology and Botany, Hungarian Academy of Sciences, H-2163 Vácrátót, Alkotmány u. 2-4, Hungary; *) e-mail bdz@botanika.hu*

Abstract: Due to the long tradition of the Braun-Blanquet approach, many relevés using this approach have been made. Recent developments in vegetation-plot databases provide an opportunity to effectively use these relevés to study ecological problems as well. Opinions differ, however, concerning the applicability of these datasets, often with their use being restricted to exploration and hypothesis generation only.

We assert that preferential sampling, which is characteristic of the Braun-Blanquet approach, means using a special definition of statistical population rather than non-random sampling. We present a case study, where consequences of using a preferential and non-preferential definition of statistical population are studied. Although the traits of stands that are preferred or avoided by the phytosociologist during preferential sampling can be identified, there are no general rules that could predict the difference between the preferential and non-preferential datasets obtained for the same object.

Keywords: Braun-Blanquet approach, Phytosociological database, Sampling criteria, Statistics

INTRODUCTION

In western and Central Europe there is a long tradition to collect vegetation data according to the Braun-Blanquet approach (WESTHOFF & VAN DER MAAREL 1980). As for logical context, phytosociology has remained a minor sister of plant taxonomy (BRAUN-BLANQUET 1928). The idea that associations are biological units that can be typified and described in analogy to species (PIGNATTI et al. 1997, but see differences in MIRKIN 1989) was developed in the pioneer period at the beginning of 20th century. Therefore, similarly to idiotaxonomy, where instead of within-species variation, properties of typical individuals are considered when describing species, phytosociologists collected data from the so-called typical stands (WESTHOFF & VAN DER MAAREL 1980) and a huge number of relevés have been made (RODWELL 1995, PIGNATTI et al. 1997). Recent developments of vegetation-plot databases (BRISSE et al. 1995, FONT & NINOT 1995, MUCINA et al. 2000, EWALD 2001, HENNEKENS & SCHAMINÉE 2001, CHYTRÝ & RAFAJOVÁ 2003) provide an opportunity not only for effective use of these relevés in the unification of vegetation classification on a European scale (MUCINA et al. 1993, RODWELL et al. 2002), but for studying ecological problems as well (e.g. KIENAST et al. 1998, DUCKWORTH et al. 2000, CHYTRÝ et al. 2003). These large datasets are useful tools for ecological research (especially in studies of long-term vegetation changes; see e.g. RUPRECHT & BOTTA-DUKÁT 1999, HOLEKSA & WOZNIAK 2005), even if there may be inconsistencies in concepts and methods of sampling within the datasets, which may yield artefacts (cf. CHYTRÝ 2001), and some areas and vegetation types may be over- or underrepresented (cf. CHYTRÝ & RAFAJOVÁ 2003).

Recently, LÁJER (2007), in his reflection on the paper of RÉDEI et al. (2003), argues that ecological hypotheses cannot be tested in datasets consisting of relevés made according to the Braun-Blanquet approach. If Lájer's opinion is true, applicability of phytosociological datasets is strongly restricted; i.e., they can be used only for exploration and hypothesis generation.

In the first part of this paper we consider this question from a theoretical point of view and claim that preferential sampling means creating a special definition of statistical population. In the second part of the paper we present a case study, where the consequences of using a preferential and non-preferential definition of statistical population were studied.

THEORETICAL CONSIDERATIONS

Population *vs.* sample

Statistical analysis is used when we want to draw a conclusion about a set(s) of things without measuring all of its (their) elements. The set, which is the subject of our statistical analysis, is called (statistical) universe or population (ZAR 1999). Note that in this paper we use the term population not in the biological, but in the statistical sense. Measuring all elements of a population is impossible if its size is infinite (e.g. an experiment can be repeated unlimited times) or unpractical if its size is finite, but large (e.g. measuring all "individuals" in a site). Therefore, researchers choose a subset of the population – the sample – and measure its elements only (ZAR 1999), but they extend the results to the whole population.

The first step of any study operating with statistical analysis is defining the population. Population may be any set, but it has to be defined exactly according to the purpose of the study before sampling. An exact definition of population is essential because results will be valid for and only for the population defined.

Although definition of the studied population is very important, it is often not exact enough in publications. For example, RÉDEI et al. (2003) used terms "open dolomite grassland" and "*Seseli leucospermi-Festucetum pallentis*" as synonym names of the same population. Open dolomite grassland can be defined by dolomite substrate and a threshold of maximal vegetation cover. On the other hand, *Seseli leucospermi-Festucetum pallentis* is defined by its characteristic species composition, irrespective of abiotic conditions including substrate (sometimes it can be found on limestone; KUN & ITTZÉS 1995). In this special case, the association (i.e., *Seseli leucospermi-Festucetum pallentis*) is a wider category than the corresponding habitat category (i.e., open dolomite grassland), but often the opposite is true. Let us consider the following simple hypothetical situation. Someone wants to analyze the herb layer of a sessile oak-turkey oak forest. He/she can find in the phytosociological literature (e.g. BORHIDI 2003) that phytosociologists call this vegetation type *Quercetum petraeae-cerris*. If he/she collects and analyzes relevés from this association, results will not be valid for sessile oak-turkey oak forests, because most phytosociologists made relevés only in the old-growth forests, where the herb layer differs from the understorey vegetation of other regrowth phases (CSONTOS 1996).

Random sampling

After defining the population, the next step is sampling, i.e., choosing the elements of the sample. Standard statistical procedures give valid results only if:

(1) all elements of the population are chosen with equal probability;

(2) elements of the sample are independent from each other, that is choosing one element does not influence the probability of choosing any other elements.

If these conditions are not satisfied, sample characteristics (e.g. mean) will not be an unbiased estimation of population characteristics (e.g. expected value). While population characteristics are fixed values, their estimates are random variates. In the standard statistical procedures, distribution of these estimates (including test statistics) can be calculated from the distribution of sample elements, if and only if the elements are independent of each other.

Random sampling (i.e., choosing elements of the sample randomly) is suggested to fulfill both preconditions. If the precondition of independence cannot be satisfied – e.g. properties of related species are similar due to their shared phylogenesis; or similarity of quadrats increases with their decreasing topographical distance (spatial autocorrelation) – special statistical methods (spatial statistics, phylogenetic correction methods, randomization techniques, mixed models etc.) are required. Here we do not consider these special cases, because phytosociological relevés are usually analyzed using standard statistical methods.

There are sampling techniques (e.g. using random co-ordinates) that guarantee random sampling. However, in field studies they are rarely used, and researchers usually allocate plots of relevés without formal selection.

What does preferential sampling mean in phytosociology?

In theory, preferential sampling means that researchers make relevés only in the most typical stands. Our personal experience, the literature and consultations with Hungarian phytosociologists suggest that in practice it means defining a population according to the rules of this school (i.e., the studied population does not contain the so-called “atypical”, “non-homogeneous”, “transitional”, species-poor, disturbed etc. stands; WESTHOFF & VAN DER MAAREL 1980, KENT & COKER 1992, DIERSCHKE 1994), and sampling within this population without further selection.

Formal and conventional criteria of statistical tests

In statistics there are strict formal rules of using tests. For example, analysis of variance (ANOVA) requires the normality, additivity, homoscedasticity and independence of errors (e.g. ZAR 1999). If these criteria are not satisfied, the estimation of Type I error rate will be biased. From a pure mathematical point of view, there is no difference between a small and large bias of estimation. In practice, however, we often consider only the Type I error relative to the predefined treshold (significance level), thus a small bias is tolerable.

In practice, therefore, we use less strict criteria instead of formal ones. We call these less strict citeria conventional criteria, because they were gradually applied and accepted in the practice of science, and are used by convention. For example, ANOVA estimates the probability of Type I error with small bias, if error has symetric but not normal distribution. Thus we can use ANOVA by convention, even if the error terms do not strictly follow normal distribution.

We agree with LÁJER (2007) that phytosociological sampling does not satisfy the formal criterion of randomness. We do not agree with his conclusion, however, that statistical methods cannot be applied to analyze phytosociological data.

There are two types of statistical analysis: exploration and hypothesis testing (e.g. HALLGREN et al. 1999). The two types of analysis differ in their aims, but often use the same methods. At least for exploration, statistical tests can be applied, even if their formal criteria are not satisfied. In this case the calculated Type I error rates are interpreted as descriptive statistics instead of probability.

Of course, conventional criteria in phytosociology should rule out the careless use of statistics, but permit the researcher to carefully decide which formal criteria of which procedure should be applied less rigorously. Application of conventional criteria, however, should couple with moderate interpretation of results.

What are the conventional criteria of using phytosociological data to compare different habitats at a regional or sub-continental scale as is the case of studies like that of RÉDEI et al. (2003)?

(a) The relevés should represent the abiotic conditions and biogeographical context of the study area rather than impacts of local disturbance (unless disturbance is well defined and widespread, e.g. mowing, grazing, flood effect, etc.).

(b) The location of the sampling units in each study site should be selected by the same and reproducible algorithm. This is one of the focal points in our debate with LÁJER (2007), because the algorithm of selection is not formally defined. In fact, during the selection procedure, the surveyor makes a general picture of the vegetation of the study site and compares it with his or her previous knowledge of the associations recognized. Then, the surveyor chooses the most “characteristic” stands that represent the recognized associations as well as the local deviations from the typical compositions of those associations. This algorithm is not independent of the persons who apply it, but according to the Central-European phytosociological practice, the surveyors can learn and reproduce this algorithm effectively. Thus, the general criterion of “objectivity” of natural sciences is not strictly fulfilled but the reproducibility, at least among trained phytosociologists (KENT & COKER 1992, MUCINA 1997), is fulfilled.

(c) All the relevés should be taken using the same method considering quadrat size, abundance estimation, taxonomical system of species, etc.

We take the above three points into account together in order to judge the selection of study sites by RÉDEI et al. (2003), i.e., each type of bedrock represented by one separate region. Finally, we consider the sampling and the statistical procedures used by RÉDEI et al. (2003) acceptable regarding the conventional rules of phytosociology.

Concluding remarks

In our opinion, even though phytosociological sampling does not satisfy the formal criteria of statistical analysis, phytosociological relevés can be analyzed using statistical methods if conventional criteria are fulfilled. Unfortunately, statistical handbooks for biologists (e.g. ZAR 1999, SOKAL & ROHLF 1981) treat formal criteria only, and conventional criteria are neglected in these books and in phytosociological literature.

Rules of interpretation are the most important conventional criteria. Which population is represented by the phytosociological relevés as sample and to which the results can be extended? We think that it is not a simple question to which a generally valid answer exists. Rather, the researcher using data from literature or databases should carefully consider the preferences of phytosociologists who made the relevés in each case. In most cases, we can have good guess, but this cannot be verified unless asking the phytosociologist himself. In the next part of the paper we present a case study where we tested our *a priori* expectations of the differences of results based on random and preferential sampling of the “same” object.

CASE STUDY: COMPARISON OF PREFERENTIAL AND NON-PREFERENTIAL SAMPLING

It is generally accepted that phytosociology of Zürich-Montpellier school prefers species-rich stands (c.f. CHYTRÝ 2001), especially those that are rich in characteristic species (specialists, rare species, narow-range endemics or indicators of special phytogeographic influence), and avoids sampling transitional and degraded stands (WESTHOFF & VAN DER MAAREL 1980). Symptoms of degradation may include the decrease of species richness and the occurrence of ruderal strategy species (weeds). Therefore, in this case study such vegetation characteristics that are sensitive to this preference were used; i.e., species richness, proportion of species groups according to geographic distribution, phytosociological preference, and Raunkiaer life form of species.

We expect that preferential sampling overestimates species richness, proportion of specialists, narrow-range endemics and indicators of special phytogeographic influence, and underestimates the proportion of weeds and generalist species with wide distribution.

Beyond the characterization of vegetation within one site the usual aim of the analysis is to compare two or several sites. From a statistical point of view, the former aim is estimation while the latter one is hypothesis testing. Small differences between preferential and non-preferential sampling in the estimation can be enlarged in hypothesis testing. This can result in significant differences between the conclusions based on the different sampling designs. We will provide an example of this effect. Conclusions of between-site comparisons, however, can be the same, even if significant within-site differences between sampling methods exist. We, therefore, also compared conclusions of between-site comparisons based on the two sampling methods.

Because traditional phytosociologists search for typical stands rather than variation, we expected that if we took samples in two sites within the same vegetation type, preferential sampling will emphasize the general properties of the vegetation type, and consequently underestimate the differences between sites.

Study areas and the sampling process

The sample areas Csévharaszt and Fülöpháza are located on the calcareous sand deposit of the Danube river, the hilly region of the Danube-Tisza interfluve (part of the Hungarian Plain). The natural vegetation is sand forest-steppe, a mosaic composed of patches of open sand grasslands, mesic grasslands and juniper-poplar groves. A gradual increase in aridity appears throughout the region from northwest (Csévharaszt) to southeast (Fülöpháza) and accordingly, the structure of the forest steppe mosaic also changes shifting from higher to lower cover of woody vegetation. The open sand grassland is also richer in perennial, xero-mesic forest-steppe species at Csévharaszt, while the proportion of annual species of continental and submediterranean origin is higher at Fülöpháza (KOVÁCS-LÁNG et al. 2000).

A 50 ha (400 × 1250 m) sample area was selected at both sites where the open dry sand grassland (*Festucetalia vaginatae* SOÓ 1957) was sampled in parallel preferentially by the "classic" Braun-Blanquet approach (30 quadrats) and non-preferentially using 40–50 randomly located quadrats, of which 30 also were randomly selected for comparison. Plot size was 4 × 4 m in both sampling designs. Preferential sampling was done by the same person at both sites.

Attributes used for comparison

According to their geographic distribution (HORVÁTH et al. 1995) species were divided into four groups: (1) Endemic and subendemic species with small geographical range (endemics); (2) Species with submediterranean range (submediterranean); (3) Species with continental range (continental); (4) Species with wide geographic distribution (widely distributed).

To characterize the phytosociological preference of the species the next four categories, based on BORHIDI (1995), were applied: (1) Specialists of calcareous open sand grasslands (sand grassland species); (2) Xero-mesic species of sand forest-steppes (forest-steppe species); (3) Generalist species of dry grasslands (dry grassland species); (4) Species indicating anthropogenic disturbance (weeds).

From the Raunkiaer life forms we applied the rough distinction forming only two categories according to the life span of species: annuals (Th, TH of Raunkiaer) and perennials (G, H, Ch, N, M of Raunkiaer).

Statistical analysis

To avoid bias caused by differences in cover estimation between researchers (SYKES et al. 1983, KENNEDY & ADDISON 1987, HAHN & SCHEURING 2003), only presence-absence data were used in all analyses.

The effect of sampling method and region on species composition was explored using distance-based redundancy analysis (LEGENDRE & ANDERSON 1999) using a square root of a complement of Jaccard similarity as a resemblance measure (LEGENDRE & LEGENDRE 1998, PODANI 2000). Whether the sampling method affects the difference between the two regions is addressed by looking at the partial effect of the interaction between these two factors, after accounting for the differences between the regions and between the sampling methods (using distance-based partial redundancy analysis). The probability of Type I error was determined using a permutation test. Distance-based redundancy analyses were done by CANOCO 4.52 (TER BRAAK & ŠMILAUER 1998).

Species richness and proportion of the above-mentioned species groups were analyzed using standard univariate methods. First we explored the differences between two statistical populations, i.e., non-preferential and Braun-Blanquet data sets were compared within sites. The Smirnov test was applied to decide if it is reasonable to conclude that the studied character has the same cumulative frequency distribution in the two data sets. The advantage of Smirnov test is that its null-hypothesis concerns the whole rather than some specific aspect of distribution (e.g. central tendency or dispersion). The disadvantage of this test is that it is weaker than those concerning special features of distribution (e.g. the Wilcoxon-Mann-Whitney test) (SPRENT & SMEETON 2001). For this reason medians were also compared by Wilcoxon-Mann-Whitney (WMW) test (SPRENT & SMEETON 2001). In the second step, we studied whether different sampling strategies yield different conclusions in the comparison of sites. Therefore, Wilcoxon-Mann-Whitney tests were applied to compare sites, because such data (where a normality criterion is not satisfied) are often analyzed using non-parametric methods (e.g. KRÖEL-DULAY et al. 2004). Univariate statistical tests were performed using SPSS 10 (SPSS Inc. 1999). In the case of the Smirnov tests, exact probabilities of Type I error were calculated.

RESULTS

Distance-based redundancy analysis revealed significant interaction between sites and sampling methods (F = 1.50, P = 0.003), because the random and preferential dataset significantly differed at Fülöpháza (F = 1.61, P = 0.0059) but not at Csévharaszt (F = 1.29, P = 0.0524). Only two species' frequencies differ considerably in the preferential and non-preferential dataset at Fülöpháza. *Cerastium semidecandrum* was more frequent, while *Salsola kali* was less frequent in the preferential dataset. If these two species were excluded from the analyses the difference between sampling methods became non-significant (F = 1.07, P = 0.3034).

Tables 1 and 2 summarize the results of a comparison based on species numbers and plant (vegetation) attributes. A within-site comparison of sampling method gave different results for the two sites. While the results of "preferential" and "non-preferential" sampling do not show differences at Fülöpháza, there are significant differences between the results of the two samplings at Csévharaszt in the number of species detected, the proportion of wide distributed species, as well as the proportion of sand grassland species (Table 2). As histograms show, at Csévharaszt species-poor stands (Fig. 1a), stands poor in widely distributed species (Fig. 1b), and stands rich in open sand grassland specialists (Fig. 1c) have been preferred.

In a between-sites comparison, preferential sampling did not produce significant differences in the proportion of widely distributed species, whereas, random sampling showed a significantly higher proportion of these species at Csévharaszt (Table 1). Sampling methods give different results when proportion of continental species in the two sites were compared, although there were no significant differences between sampling methods within sites (Table 2). Preferential sampling did not produce significant differences, while random sampling showed a significantly higher proportion of continental species at Fülöpháza

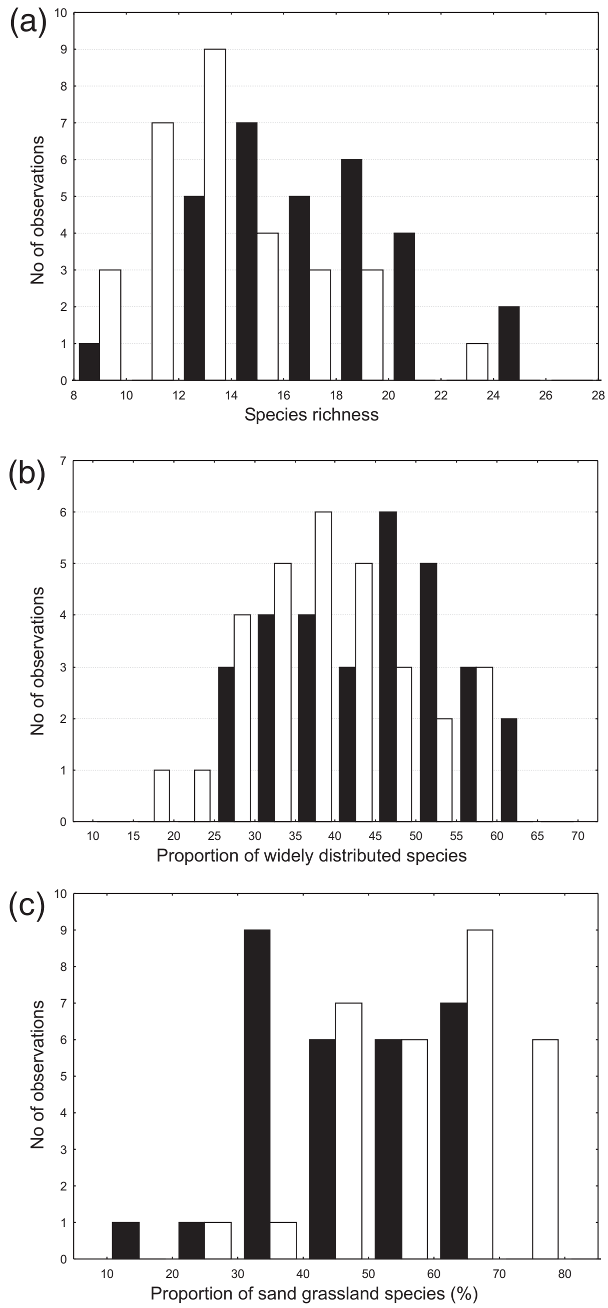

Fig. 1. (a) – Frequency distribution of species richness in the preferential (empty columns) and non-preferential (filled columns) data sets from the Csévharaszt site. (b) – Frequency distribution of the proportion of widely distributed species in the preferential (empty columns) and non-preferential (filled columns) data sets from the Csévharaszt site. (c) – Frequency distribution of the proportion of sand grassland species in the preferential (empty columns) and non-preferential (filled columns) data sets from the Csévharaszt site.

Table 1. Means of characters and probability of Type I error in the between-site comparions using the Wilcoxon-Mann-Whitney test. (1) means in Csévharaszt; (2) means in Fülöpháza; (3) probability of Type I error. Probabilities of Type I error less than 5% are marked by bold numbers.

	Preferential data set			Non-preferential data set		
	(1)	(2)	(3)	(1)	(2)	(3)
Species richness	14.3	15.9	0.071	17.5	16.1	0.177
Life form						
annuals	29.9%	39.4%	**0.037**	34.2%	43.1%	**0.016**
perennials	70.1%	60.6%	**0.037**	65.8%	56.9%	**0.016**
Geographic distribution						
endemics	27.1%	25.5%	0.424	23.9%	23.6%	0.988
submediterranean	20.6%	27.7%	**0.001**	19.0%	24.3%	**0.030**
continental	40.0%	39.5%	0.947	36.1%	42.0%	**0.010**
widely distributed	28.7%	23.6%	0.101	35.6%	27.3%	**0.003**
Habitat preference						
sand grassland species	59.1%	61.2%	0.584	48.3%	66.2%	< **0.001**
dry grassland species	29.2%	34.6%	0.078	32.1%	28.4%	0.625
forest-steppe species	8.9%	1.3%	< **0.001**	15.6%	2.1%	< **0.001**
weeds	2.8%	2.9%	0.290	4.8%	3.3%	0.718

(Table 1). The explanation of this counter-intuitive result may be that the proportion of continental species is underestimated at Fülöpháza, but overestimated at Csévharaszt by preferential sampling (Table 1). The proportion of sand grassland species does not show a significant difference between sites according to preferential, while it is significantly higher at Fülöpháza according to non-preferential sampling (Table 1).

DISCUSSION

The analyses and comparisons of data gained using the two sampling approaches produced less difference within sites than expected. The difference in species composition between sampling methods at Fülöpháza was caused by only two species. At the sampling date, recognition of these species might be ambiguous because *Cerastium* was in the senescence stage, while *Salsola* was present as a seedling. Thus, this difference is caused rather by the different researchers than the different sampling method. In the derived variables no significant difference was found between sampling methods at Fülöpháza, and only three significant differences were found at Csévharaszt.

The perennial open sand grassland has relatively large, non- or slightly disturbed stands at both sites. The species pool can be found on a rather small area, and stands show considerable resistance against weeds and invaders. After slight or medium disturbances the dry grassland species ("interior pioneers") become more abundant (RÉDEI 2005). Therefore atypical or weedy patches, which are avoided by traditional phytosociologists, are generally rare in this habitat. However, the two studied sites slightly differ in this respect. At Fülöpháza, where large, open grassland patches are dominating in the forest-steppe mosaic, the grassland, in spite of the disturbances caused by severe droughts and/or trampling, displayed rather homogenous phytosociological character of the *Festucetalia vaginatae* type. At Csévharaszt, however, landscape is strongly fragmented by intermingled patches of the natural

Table 2. Probabilities of Type I error in within-site comparisons of sampling methods. Probabilities of Type I error less than 5% are marked by bold numbers.

	Csévharaszt		Fülöpháza	
	Smirnov test	WMW test	Smirnov test	WMW test
Species richness	**0.003**	< **0.001**	0.980	0.970
Life form				
annuals	0.549	0.201	0.353	0.403
perennials	0.549	0.201	0.353	0.403
Geographic distribution				
endemics	0.320	0.120	0.688	0.496
submediterranean	0.995	0.524	0.198	0.077
continental	0.321	0.157	0.525	0.318
widely distributed	0.116	**0.027**	0.334	0.086
Habitat preference				
sand grassland species	0.060	**0.003**	0.371	0.314
dry grassland species	0.683	0.322	0.341	0.096
forest-steppe species	0.059	0.079	0.175	0.137
weeds	0.166	0.087	0.998	0.787

juniper-poplar groves and non-indigenous *Robinia* plantations. The existence of this variable mosaic produces a considerable edge effect, significantly increasing the proportion of species alien to the sand grassland community, e.g. mesic forest-steppe and weed species (KOVÁCS-LÁNG et al. 2000). During the search for typical stands these “alien for sand grassland” species were consequently avoided.

It is slightly surprising at first glance that the number of species per quadrat is higher in random samples. A possible explanation is that open sand grasslands are locally species-poor communities with a small species pool. An increasing number of species often indicates transition to stands of other, more species-rich communities (sand steppe, closed interdune grasslands, forest edge) or degradation. That is why in this case preferential sampling avoided the species-rich patches. In other communities (e.g. *Molinia* meadows) where species richness decreases due to degradation, we would probably find an opposite trend.

Preferential sampling did not magnify, as was expected, but rather hid the differences between the two sites. It should be emphasized that we compared the same vegetation type between two sites. In our opinion, if we had compared two distinguished vegetation types different due to their phytogeographic position, preferential sampling would magnify the differences.

CONCLUSIONS

From the theoretical considerations in the first part of this paper we conclude that preferential relevés made using the Braun-Blanquet approach can be analyzed by statistical tests. The populations, to which the results of the tests can be extended, however, differ from the populations of non-preferential sampling designs. In the case study we demonstrated the character, direction and significance of this difference.

Although the sorts of stands that are preferred or avoided by a phytosociologist during preferential sampling can be identified, there are no general rules that could predict the difference between the preferential and non-preferential datasets gained from the same object. The reason is that preference of the phytosociologists depends on their preconcepts about the studied vegetation. For example, if the phytosociologist wants to compare two associations in the same location, she/he will emphasize the differences. Whereas, if she/he samples the stands of the same association in different locations, she/he will emphasize the similarities. Therefore, to avoid misinterpretations before analyzing preferential datasets we have to familiarize ourselves – as far as possible – with the (pre)concepts of the phytosociologists who have collected the data.

Acknowledgement: We are grateful to Konrád Lájer, whose critique inspired us to write this paper. Our thanks goes to Sándor Bartha, Milan Chytrý, Tibor Kalapos and two anonymous referees for their helpful comments. This research was supported by the Hungarian Scientific Research Fund (OTKA T 034790) and by the National Research and Development Programme (NKFP 3B/0008/2002).

REFERENCES

BORHIDI A. (1995): Social behaviour types, the naturalness and relative ecological indicator values of the higher plants in the Hungarian Flora. *Acta Bot. Hung.* 39: 97–181.

BORHIDI A. (2003): *Plant communities of Hungary*. Akadémiai Kiadó, Budapest.

BRAUN-BLANQUET J. (1928): *Pflanzensoziologie*. Springer Verlag, Wien.

BRISSE H., DE RUFFRAY P., GRANDJOUAN G. & HOFF M. (1995): The Phytosociological Database "SOPHY" Part 1: Calibration of indicator plants, Part II: Socio-ecological classification of the relevés. *Ann. Bot. (Rome)* 53: 177–223.

CHYTRÝ M. (2001): Phytosociological data give biased estimates of species richness. *J. Veg. Sci.* 12: 439–444.

CHYTRÝ M., TICHÝ L. & ROLEČEK J. (2003): Local and regional patterns of species richness in Central European vegetation types along the pH/calcium gradient. *Folia Geobot.* 38: 429–442.

CHYTRÝ M. & RAFAJOVÁ M. (2003): Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75: 1–15.

CSONTOS P. (1996): *Az aljnövényzet változásai cseres-tölgyes erdők regenerációs szukcessziójában* (*Regeneration succession of sessile oak – turkey oak forests: Processes in the herb layer*). Scientia Publishing, Budapest.

DIERSCHKE H. (1994): *Pflanzensoziologie*. Verlag Eugen Ulmer, Stuttgart.

DUCKWORTH J.C., BUNCE R.G.H. & MALLOCH A.J.C. (2000): Vegetation gradients in Atlantic Europe: the use of existing phytosociological data in preliminary investigations on the potential effects of climate change on British vegetation. *Global Ecol. Biogeogr.* 9: 187–199.

EWALD J. (2001): Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. *Ber. Reinhold Tüxen Ges.* 13: 53–69.

FONT X. & NINOT J.-M. (1995): A regional project for drawing up inventories of flora and vegetation in Catalonia (Spain). *Ann. Bot. (Rome)* 53: 99–105.

HALLGREN E., PALMER M.W. & MILBERG P. (1999): Data diving with cross-validation: an investigation of broad-scale gradients in Swedish weed communities. *J. Ecol.* 87: 1037–1051.

HAHN I. & SCHEURING I. (2003): The effect of measurement scales on estimating vegetation cover: a computer experiment. *Community Ecol.* 4: 29–33.

HENNEKENS S.M. & SCHAMINÉE J.H.J. (2001): TURBOVEG, a comprehensive database management system for vegetation data. *J. Veg. Sci.* 12: 589–591.

HOLEKSA J. & WOZNIAK G. (2005): Biased vegetation patterns and detection of vegetation changes using phytosociological databeses. A case study in the forest of the Babia Góra National Park (the West Carpathians, Poland). *Phytocoenologia* 35: 1–18.

HORVÁTH F., DOBOLYI Z.K., MORSCHHAUSER T., LŐKÖS L., KARAS L. & SZERDAHELYI T. (1995): *FLÓRA database 1.2*. MTA-ÖBKI & MTM Növénytára, Vácrátót.

KENNEDY K.A. & ADDISON P.A. (1987): Some consideration for the use of visual estimates of plant cover in biomonitoring. *J. Ecol.* 75: 151–157.

KENT M. & COKER P. (1992): *Vegetation description and analysis. A practical approach*. John Wiley & Sons, Chichester.

KIENAST F., WILDI O. & BRZEZIECKI B. (1998): Potential impacts of climate change on species richness in mountainforests – An ecological risk assessment. *Biol. Conservation* 83: 291–305.

KOVÁCS-LÁNG E., KRÖEL-DULAY GY., KERTÉSZ M., FEKETE G., BARTHA S., MIKA J., DOBI-WANTUCH I., RÉDEI T., RAJKAI K. & HAHN I. (2000): Changes in the composition of sand grasslands along a climatic gradient in Hungary and implications for climate change. *Phytocoenologia* 30: 385–407.

KRÖEL-DULAY GY., ÓDOR P., PETERS D.P.C. & HOCHSTRASSER T. (2004): Distribution of plant species at a biome transition zone in New Mexico. *J. Veg. Sci.* 15: 531–538.

KUN A. & ITTZÉS P. (1995): A *Seseli leucospermum* W. et K. és a nyílt dolomitsziklagyep (*Seseli leucospermo-Festucetum pallentis*) előfordulása szarmata mészkövön (Occurence of *Seseli leucospermum* W. et K. and *Seseli leucospermo-Festucetum pallentis* community on sarmatian limestone). *Bot. Közlem.* 82: 27–34 .

LEGENDRE P. & LEGENDRE L. (1998): *Numerical ecology*. Ed. 2. Elsevier, Amsterdam.

LEGENDRE P. & ANDERSON M.J. (1999): Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69:1–24.

MIRKIN B.M. (1989): Plant taxonomy and syntaxonomy: a comparative analysis. *Vegetatio* 82: 35–40.

MUCINA L. (1997): Classification of vegetation: Past, present and future. *J. Veg. Sci.* 8: 751–760.

MUCINA L., BREDENKAMP G.J., HOARE D.B. & MCDONALD D.J. (2000): A national vegetation database for South Africa. *S. African J. Sci.* 96: 497–498.

MUCINA L., RODWELL J. S., SCHAMINÉE J.H.J. & DIERSCHKE H. (1993): European vegetation survey: current state of some national programmes. *J. Veg. Sci.* 4: 429–439.

PIGNATTI S., DOMINICI E. & PIETROSANTI S. (1997): European vegetation survey – from the methodological discussion to the first approximation. *Ann. Bot. (Rome)* 45: 5–16.

PODANI J. (2000): *Introduction to the exploration of multivariate biological data*. Backhuys Publishers, Leiden.

RÉDEI T. (2005): *A növényi fajkészlet eloszlása nyílt szárazgyepekben* (*The distribution of plant species pool in open dry grasslands*). PhD. Thesis, Budapest.

RÉDEI T., BOTTA-DUKÁT Z., CSIKY J., KUN A. & TÓTH T. (2003): On the possible role of local effects on species richness of acidic and calcareous rock grasslands in northern Hungary. *Folia Geobot.* 38: 453–467.

RODWELL J.S. (1995): The European Vegetation Survey questionnaire: an overview of phytosociological data, vegetation survey programmes and databases in Europe. *Ann. Bot. (Rome)* 53: 87–98.

RODWELL J.S., SCHAMINÉE J.H.J., MUCINA L., PIGNATTI S., DRING J. & MOSS D. (2002): *The diversity of European vegetation. An overview of phytosociological alliances and their relationships to EUNIS habitats*. EC-LNV. Report EC-LNV. nr. 2002/054, Wageningen.

RUPRECHT E. & BOTTA-DUKÁT Z. (1999): Long-term vegetation textural changes of three fen communities near Cluj-Napoca (Romania). *Acta Bot. Hung.* 42: 263–281.

SOKAL R.R. & ROHLF F.J. (1981): *Biometry. The principles and practice of statistics in biological research*. Ed. 2. Freeman, New York.

SPRENT P. & SMEETON N.C. (2001): *Applied nonparametric statistical methods*. Chapman & Hall/CRC, Boca Raton.

SPSS Inc. (1999): *SPSS base 10.0 user's guide*. SPSS Inc., Chicago.

SYKES J.M., HORRILL A.D. & MOUNTFORD M.D. (1983): Use of visual cover assessments as quantitative estimators of some British woodland taxa. *J. Ecol.* 71: 437–450.

TER BRAAK C.J.F. & ŠMILAUER P. (1998): *CANOCO reference manual and user's guide to Canoco for Windows*. Centre for Biometry, Wageningen.

WESTHOFF V. & VAN DER MAAREL E. (1980): The Braun-Blanquet approach. In: WHITTAKER R.H. (ed.), *Classification of plant communities*, Junk, The Hague.

ZAR J.H. (1999): *Biostatistical analysis.* Ed. 4. Prentice Hall, Upper Saddle River, New Jersey.

Received 12 December 2005, revision received and accepted 20 December 2006