Journal of Vegetation Science 12: 439-444, 2001
© IAVS; Opulus Press Uppsala. Printed in Sweden

Phytosociological data give biased estimates of species richness

Chytrý, Milan

Department of Botany, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic;
Fax +420541129513; E-mail chytry@sci.muni.cz

Abstract. Large phytosociological data sets of three types of grassland and three types of forest vegetation from the Czech Republic were analysed with a focus on plot size used in phytosociological sampling and on the species-area relationship. The data sets included 12975 relevés, sampled by different authors in different parts of the country between 1922 and 1999. It was shown that in the grassland data sets, the relevés sampled before the 1960s tended to have a larger plot size than the relevés made later on. No temporal variation in plot sizes used was detected in forest relevés.

Species-area curves fitted to the data showed unnatural shapes, with levelling-off or even decrease in plot sizes higher than average. This distortion is explained by the subjective, preferential method of field sampling used in phytosociology. When making relevés in species-poor vegetation, researchers probably tend to use larger plots in order to include more species. The reason for this may be that a higher number of species gives a higher probability of including presumed diagnostic species, so that the relevé can be more easily classified in the Braun-Blanquet classification system. This attitude of phytosociologists has at least two consequences: (1) in phytosociological data bases species-poor vegetation types are underrepresented or relevés are artificially biased towards higher species richness; (2) the suitability of phytosociological data for species richness estimation is severely limited.

Keywords: Data quality; Plot size; Relevé; Species-area curve; Vegetation data base.

Introduction

The strong tradition of Central European phytosociology, which has been developing since the early 20th century (Braun-Blanquet 1928), has led to the accumulation of a huge amount of field data in the form of vegetation relevés. Nowadays, there are certainly over 1 million relevés available in Europe, many of them being computerized in electronic data bases (e.g. Rodwell 1995; Bruelheide 2000). These data have been mostly sampled for the purposes of vegetation classification, but often they are also used for estimating species richness of plant communities, both in various descriptive papers on local vegetation typologies and in more general studies (e.g., Hobohm & Härdtle 1997; Kienast et al. 1998).

Kenkel et al. (1989) made it clear that the sampling procedure in vegetation science may differ according to whether its objective is parameter estimation (such as measuring species richness) or pattern detection (such as vegetation classification). Random arrangement of plots is necessary for parameter estimation, although systematic or stratified arrangements (Podani 1984) are considered satisfactory by many authors. However, the primary purpose of phytosociological relevés is pattern detection, i.e. describing a variety of plant communities. Therefore, they are usually sampled preferentially (Podani 1984), with plots located subjectively in the field to cover different vegetation types. Quite often, phytosociological plots are intentionally placed in vegetation stands which fit best to the researcher's *a priori* idea of vegetation type, whereas other stands are not considered (e.g. Frey 1994). If the idea of a vegetation type includes its species richness, researchers may tend to prefer sampling in either species-rich or species-poor stands. Consequently the species richness estimates from these data may be severely biased.

In this paper, I will test the hypothesis that existing phytosociological data sets are biased towards higher species richness. This may be because many researchers feel that in vegetation classification based on floristic composition, more species per relevé provide more classification criteria, thus making the classification more robust and the assignment of relevés to vegetation types more reliable.

Methods

The analysis was performed with the relevés from the Czech National Phytosociological Database (Chytrý 1997). The database currently contains approximately 40 000 relevés by different authors from different areas and phytogeographical districts of the Czech Republic. It includes both published and unpublished relevés, sampled between 1922 and 1999. Six data sets were selected from the database on the basis of the original author's assignment of relevés to phytosociological classes or orders:

A - meadows (*Molinio-Arrhenatheretea*),
B - dry grasslands (*Festuco-Brometea*),
C - reed-beds and tall-sedge vegetation (*Phragmito-Magnocaricetea*),
D - eutrophic broad-leaved forests (*Fagetalia sylvaticae*),
E - thermophilous oak forests (*Quercetalia pubescenti-petraeae*),
F - coniferous forests (*Vaccinio-Piceetea*).

The delimitation of the syntaxa followed Moravec et al. (1995). Only vascular plant records were used for the analysis, as the quality of recording cryptogams varied strongly among the authors, and in many relevés, cryptogams were not recorded at all. These data sets were further analysed separately; their details are summarized in Table 1.

As the species richness is a function of plot size, the analysis first focused on the pattern of plot sizes in the data sets. Although the followers of the Braun-Blanquet approach have achieved a rough standardization of the phytosociological plot size within broad vegetation types (e.g. Westhoff & van der Maarel 1978; Dierschke 1994), considerable variation is still evident in all the data sets. Some authors (e.g. Dierssen 1990; Hobohm 1994) have noticed that earlier generations of phytosociologists generally used larger plot sizes than those which have become standard during recent decades, at least in some vegetation types. To check for such a trend, plots sizes were plotted against sampling dates and smoother curves were fitted using the SPSS package (Anon. 1998; local linear regression with 50% points fitted). Larger plot sizes in older relevés were actually detected in the data sets A-C. Consequently, all the relevés sampled before 1970 were excluded from these data sets prior to further analysis (Table 1). This procedure should guarantee that no tacit practices specific to earlier generations of phytosociologists can affect the shape of the species-area curves.

The species richness pattern in the data sets was investigated by fitting species-area curves in semi-logarithmic space for each data set (SPSS; local linear regression with 60% points fitted).

Results

Temporal changes in the plot sizes used for making relevés showed different patterns for grassland and forest vegetation (Fig. 1). In all grassland data sets (A-C), one can recognize a clear tendency towards larger plot sizes in older relevés. Since the 1960s, however, there appears to be a collective agreement as to the use of smaller plots of ca. 16 to 25 m^2. In the forest data sets (D-F) the plot sizes appear to be roughly constant over the whole period of phytosociological research. Plots of 100-400 m^2 are clearly the most frequently used.

The species-area curves fitted by smoothing showed quite unexpected shapes (Fig. 2). The curves for meadows (A) and reed-beds and tall-sedge vegetation (C) showed a unimodal pattern. The shape of the curve for eutrophic broad-leaved forests (D) was also very close to unimodal. Thermophilous oak forests (E) showed a slight increase in species number up to the plots of 300 m^2, but no increasing trend in larger plots. Coniferous forests (F) exhibited a roughly constant species richness over the entire range of sample plots from 10 to 3000 m^2. Only the species-area curve of the dry grassland data set (B) is monotonously increasing, but the rate of increase is conspicuously lower for plots larger than 25 m^2.

Discussion

The shapes of all the species-area curves fitted to the phytosociological data considerably deviate from the standard curves reported in the literature (Palmer & White 1994; Rosenzweig 1995). In particular, unimodal species-area curves cannot correspond to any pattern found in nature, because the number of species cannot decrease with increasing area. Species richness detected from phytosociological data bases therefore reflects sampling artefacts rather than natural patterns.

Table 1. Structure of the data sets. Original data sets were used for the analysis of temporal changes in plot sizes used for making relevés. In D-F, the original data sets were also used for fitting the species-area curves. In A-C, the species-area curve fitting was based on reduced data sets containing only the relevés from 1970-1999.

	Original data sets			Reduced data sets 1970-1999	
	Nr. of relevés	Nr. of authors	Sampling dates	Nr. of relevés	Nr. of authors
A - meadows	4125	59	1931-1999	3342	51
B - dry grasslands	3262	51	1922-1999	2830	42
C - reed-beds and tall-sedge vegetation	1740	35	1932-1998	1482	30
D - eutrophic broad-leaved forests	2735	52	1924-1999	–	–
E - thermophilous oak forests	432	28	1927-1998	–	–
F - coniferous forests	681	36	1945-1999	–	–

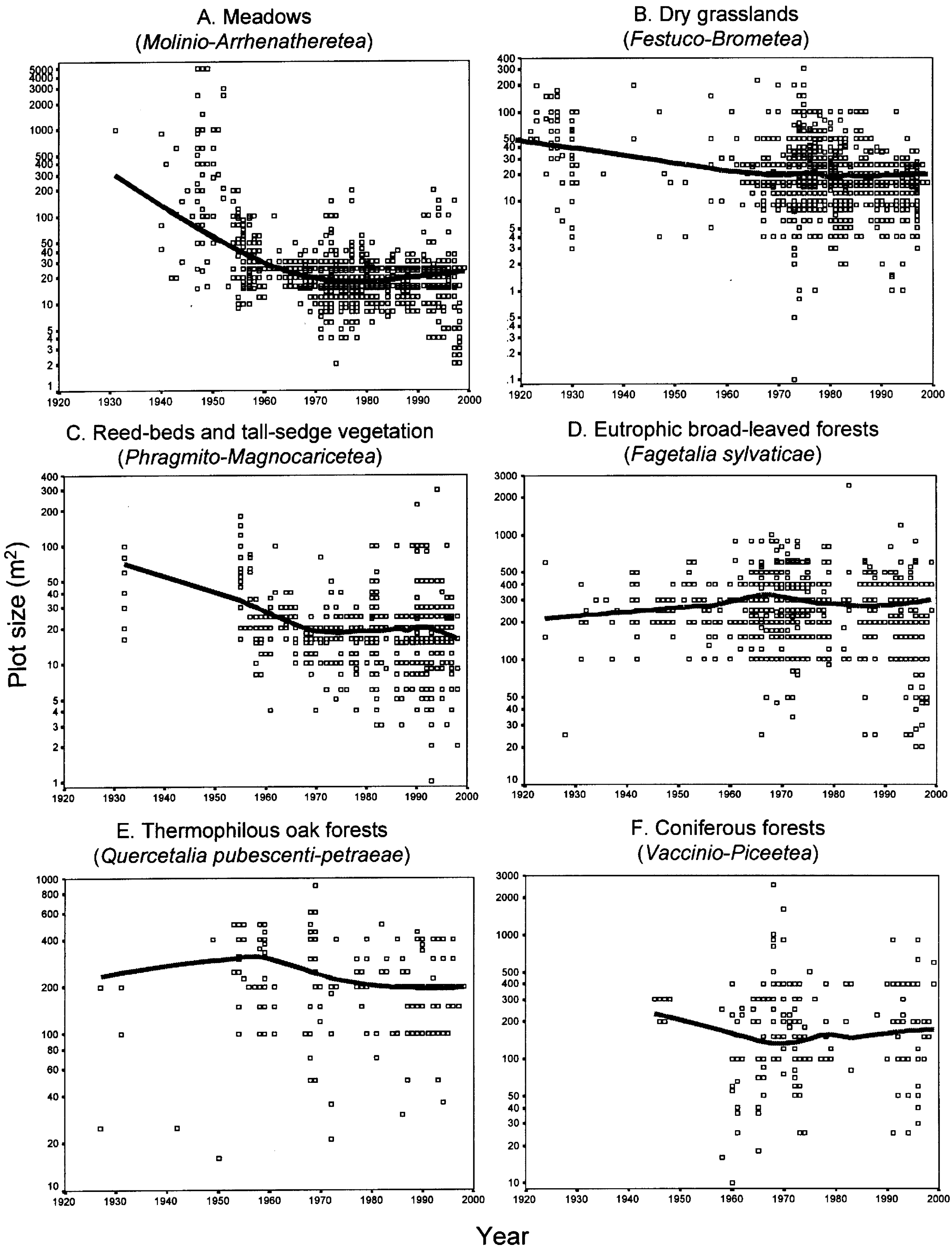

Fig. 1. Changes in the relevé plot size over time. Note the logarithmic scale on the vertical axis. Smoother curves are fitted by local linear regression.

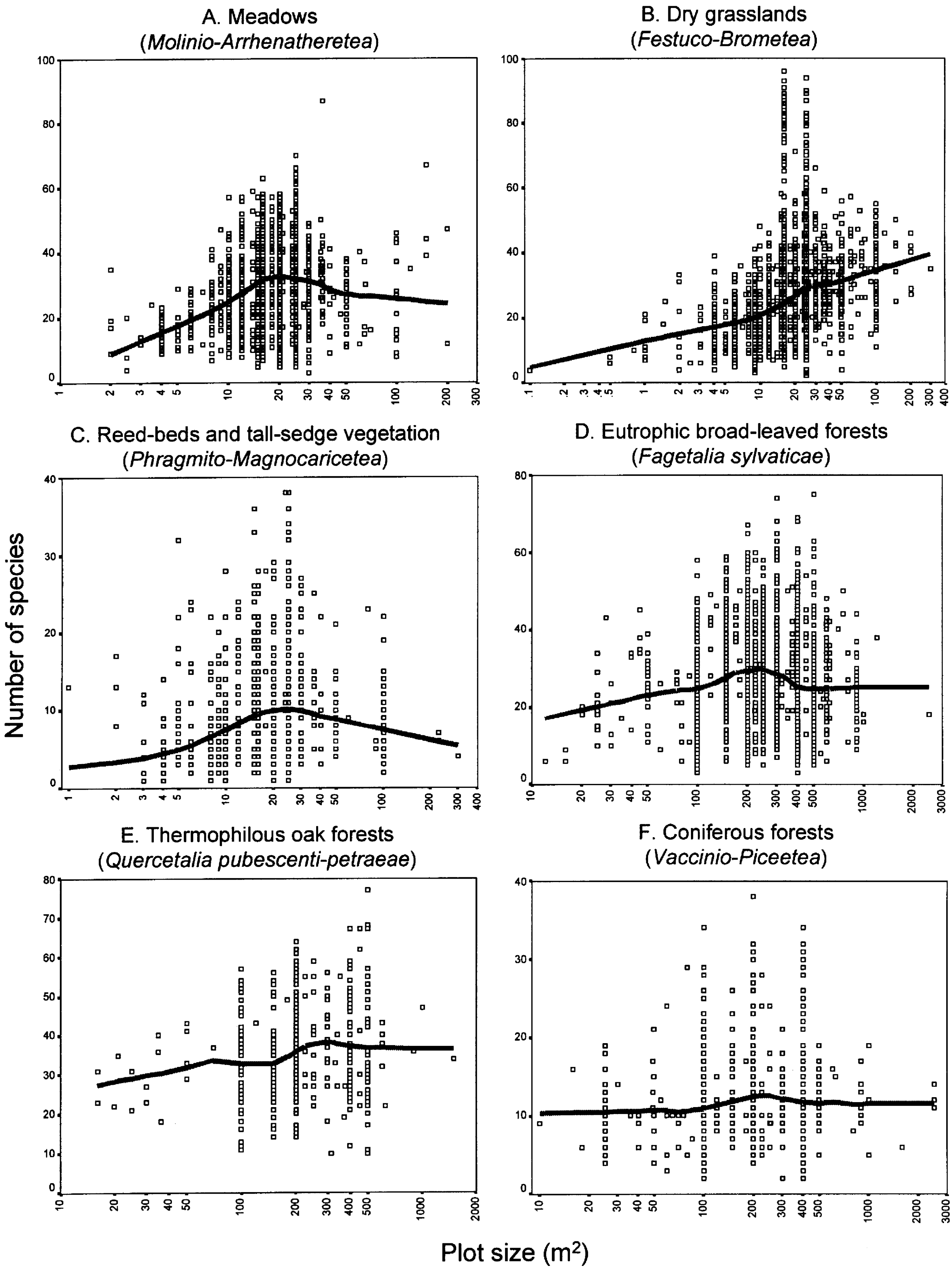

Fig. 2. Relationships between the number of vascular plant species and the relevé plot size in a semi-logarithmic space. Curves are fitted by local linear regression.

The most striking flaw is the slower rate of increase or even decrease of the curves beyond the average plot size, i.e. ca. 16-20 m^2 in grasslands and 200-300 m^2 in forests. On the whole, three possible explanations for this fact can be considered:

1. Larger plots were preferred by those researchers or research groups whose floristic lists were prepared less carefully. Such an argument could be applied for older relevés. If the earlier authors tended to use larger plots, old relevés would be over-represented among large plots. This would imply that the non-increasing right-hand parts of the curves could result from a less careful sampling practised by early phytosociologists. This explanation, however, need not be considered, as the relevés made before 1970 were removed from all the data sets in which the older relevés tended to be larger.

2. More species are overlooked when the plots become too impractically large to study carefully. This is a well-known artefact of field sampling, but it probably affects equally the 'standard-shaped' species-area curves published in the literature and the apparently flawed curves detected in this study. Quite clearly, this explanation fails in the case of unimodal curves.

3. The most probable explanation is that many phytosociologists tend to use larger plots when sampling species-poor vegetation in order to obtain species-richer relevés which they believe are better for classification. This reasoning is related to the concept of classification based on the floristic composition and notably on the diagnostic (character, differential) species. The more species there are in a relevé, the higher the probability that a sufficient number of diagnostic species will be found, and the easier and more unambiguous the assignment of the relevé to a vegetation type. This behaviour of researchers is in accordance with the common phytosociological practice of neglecting species-poor vegetation stands, though these stands can cover a major part of the landscape. If the relevés of species-poor stands are present in the data sets, they are often excluded prior to, or during the analysis (but see Kopecký & Hejný 1978 for an alternative approach). Keeping in mind that the species-poor relevés will probably be useless for further analysis, phytosociologists working in the field either entirely avoid sampling species-poor stands or enlarge the plot to 'make the stand richer'.

The traditions and methods of phytosociology in the Czech Republic have been very much the same as in Germany and the other Central European countries. Still, it remains to be tested how far the relationships detected in the Czech phytosociological data are valid in other countries. If the subjective choice of larger plots in species-poor vegetation were a more common practice among the international phytosociological community, it would present a general problem for formalized vegetation classification of currently available large data sets. Existing data on species-poor vegetation types are not only scarcer than the data on species-rich vegetation, but they may also be severely biased towards higher species richness. Therefore, estimation of species richness from phytosociological data should be limited to most frequently used plot sizes and even then, estimated values should be interpreted with caution.

Acknowledgements. I thank Leoš Klimeš, Jari Oksanen and two anonymous reviewers for critically reading the earlier version of the manuscript. The research was funded from the grants GA ČR 206/99/1523 and MSM 143100010.

References

Anon. 1998. *SPSS^R Base 8.0.* SPSS Inc., Chicago, IL.

Braun-Blanquet, J. 1928. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Springer, Berlin.

Bruelheide, H. 2000. A new measure of fidelity and its application to defining species groups. *J. Veg. Sci*. 11: 167-178.

Chytrý, M. 1997. Česká národní fytocenologická databáze: počáteční stav a perspektivy (Czech national phytosociological database: initial state and perspectives). *Zpr. Česk. Bot. Společ., Mater.* 15: 27-40.

Dierschke, H. 1994. *Pflanzensoziologie.* Ulmer, Stuttgart.

Dierssen, K. 1990. *Einführung in die Pflanzensoziologie (Vegetationskunde).* Wissenschaftliche Buchgesellschaft, Darmstadt.

Frey, H.-U. 1995. Waldgesellschaften und Waldstandorte im St. Galler Berggebiet. *Veröff. Geobot. Inst. Eidg. Tech. Hochsch. Stift. Rübel Zür*. 126: 1-280.

Hobohm, C. 1994. Einige wissenschaftstheoretische Überlegungen zur Pflanzensoziologie. *Tuexenia* 14: 3-16.

Hobohm, C. & Härdtle, W. 1997. Zur Bedeutung einiger ökologischer Parameter für die Artenvielfalt innerhalb von Pflanzengesellschaften Mitteleuropas. *Tuexenia* 17: 19-52.

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

Kienast, F., Wildi, O. & Brzeziecki, B. 1998. Potential impacts of climate change on species richness in mountain forests – An ecological risk assessment. *Biol. Conserv.* 83: 291-305.

Moravec, J., Balátová-Tuláčková, E., Blažková, D., Hadač, E., Hejný, S., Husák, Š., Jeník, J., Kolbek, J., Krahulec, F., Kropáč, Z., Neuhäusl, R., Rybníček, K., Řehořek, V. & Vicherek, J. 1995. Rostlinná společenstva České republiky a jejich ohrožení [Red list of plant communities of the Czech Republic and their endangerment]. 2nd. ed. *Severočes. Přír.*, suppl. 1995: 1-206.

Palmer, M.W. & White, P.S. 1994. Scale dependence and the species-area relationship. *Am. Nat.* 144: 717-740.
Podani, J. 1984. Spatial processes in the analysis of vegetation: theory and review. *Acta Bot. Hung.* 30: 403-425.
Rodwell, J.S. 1995. The European Vegetation Survey questionnaire: an overview of phytosociological data, vegetation survey programmes and databases in Europe. *Ann. Bot. (Roma)* 53: 87-98.
Rosenzweig, M.L. 1995. *Species diversity in space and time.* Cambridge University Press, Cambridge.
Westhoff, V. & van der Maarel, E. 1978. The Braun-Blanquet approach. In: Whittaker, R.H. (ed.) *Classification of plant communities*, pp. 289-399. W. Junk, The Hague.

Received 15 November 1999;
Revision received 5 September 2000;
Accepted 5 September 2000.
Coordinating Editor: J. Oksanen.