SAMPLING DESIGN IN LARGE-SCALE VEGETATION STUDIES: DO NOT SACRIFICE ECOLOGICAL THINKING TO STATISTICAL PURISM!

Jan Roleček[1], Milan Chytrý[1], Michal Hájek[1,2], Samuel Lvončík[1] & Lubomír Tichý[1]

1) Department of Botany and Zoology, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; e-mail honza.rolecek@centrum.cz, chytry@sci.muni.cz, hajek@sci.muni.cz, lvonca@seznam.cz, tichy@sci.muni.cz
2) Institute of Botany, Academy of Sciences of the Czech Republic, Poříčí 3a, CZ-603 00 Brno, Czech Republic

Abstract: Most of the historical phytosociological data on vegetation composition have been sampled preferentially and thus belong to those ecological data that do not fulfill the statistical assumption of independence of observations, necessary for valid statistical testing and inference. Nevertheless, phytosociological data have been recently used for various ecological meta-analyses, especially in studies of large-scale vegetation patterns. For this reason, we focus on the comparison of preferential sampling with other sampling designs that have been recommended as more convenient alternatives from the point of view of statistical theory. We discuss that while simple random sampling, systematic sampling and stratified random sampling better meet some of the statistical assumptions, preferential sampling yields data sets that cover a broader range of vegetation variability. Moreover, today's large phytosociological databases provide huge amounts of vegetation data with unrivalled geographic extent and density. We conclude that in the near future ecologists will not be able to replace the preferentially sampled phytosociological data in large-scale studies. At the same time, phytosociological databases have to be complemented with relevés of vegetation composed mostly of common and generalist species, which are under-represented in historical data. Stratified random sampling seems to be a suitable tool for doing this. Nevertheless, a methodology and input data for stratification have to be developed to make stratified random sampling an ecologically more relevant and practical method.

Keywords: Ecological methodology, Large-scale vegetation patterns, Macroecology, Phytosociology, Spatial scale, Statistical testing, Vegetation databases

INTRODUCTION

LÁJER (2007) has stirred the debate about the usability of phytosociological data in ecological studies. He emphasized the importance of independence of observations in ecological data sets if those sets are subjected to classical statistical tests. His message pertains to any kind of ecological field data. Nevertheless, in our brief discussion paper, we want to focus on its implications for the use of phytosociological data (WESTHOFF & VAN DER MAAREL 1978, HENNEKENS & SCHAMINÉE 2001), as we have been working with them for some time. Although we can see various limitations of such data, we appreciate their thus far unrivalled potential for the development of our understanding of the patterns of plant species co-existence across landscapes (EWALD 2003).

Forum: Analysis of non-randomly sampled data sets in vegetation ecology

STUDIES OF VEGETATION PATTERN ON A LARGE SCALE

When studying the patterns of vegetation in a landscape or even on a larger scale, which is the traditional domain of phytosociology (recently approached also by macroecology in more general terms; BROWN 1999, BLACKBURN & GASTON 2003), one needs a large data set that faithfully represents the variability of vegetation in the study area. The larger and more diverse the studied area is, the larger and more variable data set is necessary for the adequate representation of its vegetation. This aspect of data representativeness may be called ecological representativeness. At the same time there are theoretical requirements for the sampling procedure implied by the statistical theory that have to be followed in order to obtain valid statistical inference from the sample. This aspect refers to statistical representativeness. How do the sampling schemes of traditional phytosociology meet these multilateral requirements?

PREFERENTIAL SAMPLING

The traditional phytosociological approach is preferential sampling (BRAUN-BLANQUET 1964, MOORE et al. 1970, PODANI 1984, KENKEL et al. 1989, KENT & COKER 1992). It starts with the quick screening of vegetation variability in the landscape or in the particular locality, during which vegetation types are delimited in a researcher's mind. Such preliminary classification may be considered as an informal way of stratifying the statistical population. All distinguished vegetation types or just the vegetation types of interest are then documented using vegetation relevés that are placed subjectively, either without intentional emphasis on some species combinations or in what some phytosociologists call "typical" stands. "Typical" is a tricky term that has caused a lot of harm to phytosociology, because it has been commonly left undefined and different researchers have used it to denote different things. Probably in most cases the term was used either for the spatially prevailing vegetation or for the vegetation concentrating habitat specialists, i.e., the species differentiating one vegetation type from another.

The clear advantage of preferential sampling is that it tracks and samples nearly the full range of floristic variation in vegetation of the study area, including the rare types. It satisfies the requirement for the representation of maximum vegetation variability in the sample, while the survey resources are not wasted for the over-sampling of the prevailing vegetation types. The crucial reservation against the preferential sampling is that the structure of the sample is to an unknown extent influenced by the subjective decisions of the researcher. Some states of vegetation ("untypical" stands), though common, may be tacitly avoided (CHYTRÝ 2001, HOLEKSA & WOŻNIAK 2005).

SIMPLE RANDOM SAMPLING AND SYSTEMATIC SAMPLING

These sampling schemes have been rarely applied in phytosociology, partly for historical, but mainly for practical reasons. In simple random sampling, the rule of randomness is strictly applied so that each member of the statistical population (each site) has an equal probability of being selected (COCHRAN 1977). Systematic sampling starts with a vegetation plot placed at a randomly selected site and continues with plots located on transect or in a grid at a preset
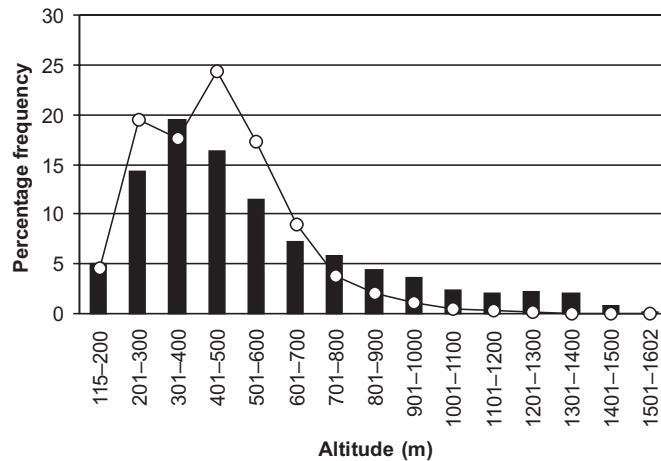
Fig. 1. A comparison of the relative distribution of 54 310 phytosociological relevés stored in the Czech National Phytosociological Database along the altitudinal gradient (bars) and the relative area of land found at altitude classes in the Czech Republic (circles). Relevés were sampled preferentially. If they were sampled randomly or systematically in geographic space, their distribution would closely fit the distribution of altitudes. Clearly, there are fewer preferentially sampled relevés than would be randomly/systematically sampled relevés at the most frequent altitudes, i.e., 200–700 m. In contrast, preferentially sampled relevés are over-represented at altitudes above 700 m, which are relatively rare in the country. This makes the database less representative in a geographic sense, but more representative in an ecological sense. Reproduced from CHYTRÝ & RAFAJOVÁ (2003), with permission of Preslia.

constant distance. Both approaches minimize the engagement of the researcher in the selection of the sampling plots and thus seemingly guarantee the statistically desirable independence. If there was no spatial autocorrelation (LEGENDRE 1993, KOENIG 1999, LEGENDRE et al. 2002, RICOTTA 2007) and periodical vegetation patterns (in case of systematic sampling; KENT & COKER 1992), independent data could really be sampled this way. However, even if statistical independence was achieved, it would not be enough, because it does not guarantee ecological representativeness of the data (MOORE et al. 1970), which is the basic condition necessary for obtaining ecologically relevant results.

AN EXAMPLE: COMPARISON OF PREFERENTIAL AND RANDOM SAMPLING

Distribution of 54 310 phytosociological relevés stored in the Czech National Phytosociological Database along the altitudinal gradient and the relative area of land found at different altitudes in the Czech Republic are compared in Fig. 1 (CHYTRÝ & RAFAJOVÁ 2003). Relevés were sampled preferentially. If they were sampled randomly or systematically in geographic space, their distribution would closely fit the distribution of altitudes. Clearly, there are fewer preferentially sampled relevés than would be randomly/systematically sampled relevés at the most frequent altitudes, i.e., 200–700 m. In contrast, preferentially sampled relevés are over-represented at altitudes above 700 m, which are relatively rare in the country. This example demonstrates that unlike simple random/systematic sampling designs,

preferential sampling does not waste resources by over-sampling sites with common values of environmental variables. Instead, it produces relatively more replicates in the environments that are rather rare in the geographic space.

In more general terms, while simple random/systematic sampling adequately represents a particular geographic space, it produces a biased sample of vegetation along ecological gradients, because these gradients are distributed non-randomly in space (MOHLER 1983, ØKLAND 2007). At the same time, ecologists are only in some cases interested in the mean properties of vegetation in a particular space, while in some other cases they look for the vegetation-environment relationships, which means they seek data sets that represent a broad range of different environmental gradients, and consequently, maximum variation in vegetation composition. Such requirements are much better met by preferential sampling than by simple random/systematic sampling schemes – in spite of the above mentioned drawbacks of the preferential sampling.

STRATIFIED RANDOM SAMPLING

Stratified random (also called restricted random) sampling is becoming more widespread in current vegetation ecology (BUNCE et al. 1996, GOEDICKEMEIER et al. 1997, STOHLGREN et al. 1997, HIRZEL & GUISAN 2002, GRABHERR et al. 2003, SMART et al. 2003). It is a compromise solution, trying to satisfy both statistician's and ecologist's demands for the representative data set. It begins with *a priori* definition of strata, i.e., habitats/vegetation types that represent particular statistical sub-populations, within which sampling sites are located randomly. The locations of sampling sites within strata may be randomly generated in geographic information system as geographic coordinates and localized in the field using GPS.

The advantage of this approach is that one can acquire variable data set with a reasonable representation of plots made in rare vegetation types. Its disadvantage is that the criteria used for stratification of the statistical population affect the resulting structure of the data set. In other words, the data sampled using stratified random design reflect mainly those environmental gradients that were used for the stratification. KNOLLOVÁ et al. (2005) explored this effect by performing classifications of subsets of phytosociological relevés from a large database, each subset being created by random resampling in different strata. They demonstrated that the resulting classifications reflected the nature of the strata used. For example, if strata were defined through climatic variables, the resulting vegetation classes mainly differed in their temperature requirements while other factors were less important for their differentiation. Because identifying the relationships between vegetation variability and environmental factors is the objective of most vegetation studies on the landscape scale, the risk of circular reasoning is obvious.

Another, more technical pitfall of stratification is that the spatial distribution of some important factors determining vegetation variability is unknown or unstable. Often it is a problem of mismatch between map resolution and ecologically relevant spatial scales: maps of environmental factors are usually available on a comparatively coarse scale, but many factors (e.g. soil moisture, soil nutrient content) vary substantially on a much finer scale. Some other environmental variables are difficult to evaluate, because the maps are based on

mean values while extreme values such as extremely dry season or extremely cold winter may be ecologically more relevant. Poorly known or hidden historical factors may also be important, as well as biotic interactions (competition, mutualism) and intrinsic factors of population dynamics. Consequently, even the most sophisticated *a priori* landscape stratification through environmental variables may fail to produce a data set which would be representative of the broad range of variation in the target vegetation of the study area.

Fine-scale stratification of the sampled area can be approached through the use of existing distribution maps of habitats or vegetation types. It may be a rather time-consuming approach, if the whole study area has to be mapped. The variation within a single broad vegetation type, however, is often the objective of the study. Then the delimitation of the vegetation type of interest at a particular locality, followed by the random selection of the exact sampling site (e.g. by an *in situ* selection of random coordinates with a GPS receiver), may be quite fast and easy. Anyway, explicit rules for the delimitation have to be formulated *a priori* otherwise the mapping units become doubtful and irreproducible. Even with such rules, maps of the same area prepared by different researchers may differ considerably.

Land cover maps based on interpreted aerial photographs or satellite images may also be efficiently used for the stratification of large areas (FANELLI et al. 2005). Unlike in vegetation maps, delimitation of strata from land cover maps is consistent across large areas and repeatable, being based on explicit rules. Still, differentiation of land cover classes may deviate from the patterns of floristic variation in plant communities, because (i) vegetation types with the same dominant species but different in species composition may have the same spectral signatures and (ii) fine-scale mosaic of different plant communities may be aggregated in the same land cover classes.

AN EXAMPLE: COMPARISON OF PREFERENTIAL AND STRATIFIED RANDOM SAMPLING

As follows from the above reasoning, preferential and stratified random sampling strategies are the only meaningful options if the aim is sampling maximum variation in the studied vegetation on a large scale. We illustrate the consequences of these two sampling strategies for the structure of vegetation data sets by comparing two independent data sets of phytosociological relevés of deciduous forests from the same region of the Czech Republic (northern Moravia and Silesia, area ca. 2100 km^2), collected using two different approaches. The first data set (Preferential) was acquired from the Czech National Phytosociological Database (CHYTRÝ & RAFAJOVÁ 2003), a source of more than 75 000 (December 2006) predominantly preferentially sampled relevés, which are frequently used for various meta-analyses (e.g. CHYTRÝ et al. 2003, 2005, HÁJEK & HÁJKOVÁ 2004, PYŠEK et al. 2005, BOTTA-DUKÁT et al. 2005, LOSOSOVÁ et al. 2006). The second data set (Strat-random) was sampled in the field in 2004–2005 using a stratified random sampling design (LVONČÍK 2006). The sampling sites were *a priori* determined by the geographic coordinates placed randomly within the strata defined by overlays of maps of environmental factors in a geographic information system. The factors used were mean annual temperature, mean annual precipitation sum, slope aspect and inclination, soil acidity and streamside location.
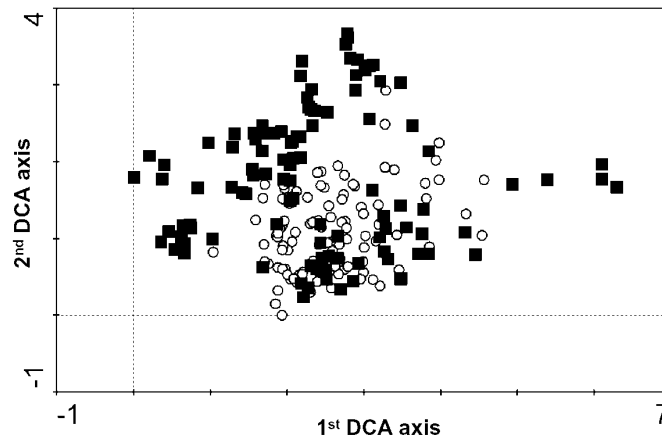
Fig. 2. DCA ordination diagram of two merged data sets of phytosociological relevés of deciduous forest vegetation made in the same region of the Czech Republic using two different sampling strategies. The preferentially sampled relevés from the Czech National Phytosociological Database (squares) and the relevés made independently by stratified random sampling (circles) represent the same broad vegetation type and both vary substantially along the first two ordination axes. While the data obtained via stratified random sampling are rather homogeneous and miss out the less abundant but ecologically important vegetation of the extremes of environmental gradients, the preferentially sampled data are ecologically heterogeneous, cover the extremes of environmental gradients rich in habitat specialists but under-represent the most common vegetation of the middle part of the gradients.

Preferential data set contained 106 relevés. Strat-random data set originally contained more relevés, but for the sake of comparison its subset of 106 randomly selected relevés was used.

Scatter plot of the detrended correspondence analysis (CANOCO program; TER BRAAK & ŠMILAUER 2002) of the merged data sets (Fig. 2) tells almost the whole story. Relevés of both data sets basically overlap and form nearly concentric clouds of points in the ordination diagram. However, relevés of Strat-random are concentrated in the centre of the ordination space, while relevés of Preferential on its periphery. It means that Strat-random is a sample of rather homogeneous vegetation, with high representation of stands composed mostly of common and generalist species. In contrast, Preferential includes many specific stands with less common species composition, typical for the extremes of environmental gradients, where habitat specialists are concentrated. Another important fact is that the main gradients of variability in species composition do correspond between the two data sets. This was confirmed also by the classification tree analysis (BREIMAN et al. 1984; not shown), which identified the Ellenberg indicator values for moisture and light as the best predictors of the variability in species composition in both data sets.

THE WAY OUT OF THE SAMPLING DILEMMA

From what was said above we conclude that no single sampling strategy can be unequivocally recommended for vegetation sampling in large-scale ecological studies. At the same time we realize that most of the data available in today's vegetation databases were sampled preferentially. Even if we developed both statistically and ecologically satisfying

design for vegetation sampling on a large scale, we cannot avoid using the preferentially sampled data in the near future. We see the solution in the plurality of approaches. In the forthcoming years, many large-scale studies will have to rely on the hitherto sampled vegetation data. Though their structure is influenced by subjective decisions of their authors (CHYTRÝ 2001, KNOLLOVÁ et al. 2005), their geographic and ecological extent and density make them irreplaceable. The first step to increase the representativeness of these data may be "filling the gaps". It rests upon the pilot analyses of the available data, identifying the geographic and ecological gaps and doing the preferential and/or stratified random sampling within these gaps (HAILA & MARGULES 1996). This practice becomes widespread in modern large-scale vegetation studies. Filling the gaps should be done in combination with stratified resampling (KNOLLOVÁ et al. 2005), performed before any analyses, as a means of increasing representativeness of the preferentially sampled data sets by reducing over-sampling of some areas or habitats. Finally, an inevitable task for the contemporary generation of vegetation scientists is the completion of the databases with the samples of "mean" or "transitory" or "untypical" vegetation, consisting mostly of generalist species. Here the stratified random sampling appears appropriate and useful. Nevertheless, its methodology and input data have to be developed to make it ecologically more relevant and practical method. We can see the following unresolved problems:

(1) Stratifications can be easily prepared for vegetation types that form coarse-grain mosaics across landscapes (e.g. forests, weed communities of arable land). Here the widely available digital maps such as elevation models, climatic models, soil types or remote sensing derived land cover can be used, because their resolution roughly corresponds to the grain of vegetation mosaics. However, this is not possible for vegetation types that form fine-scale mosaics. These include, for example, (i) vegetation complexes of wet and mesic meadows and mires, which occur in places with fine-scale variation in water availability determined by local microtopography or occurrence of springs; (ii) anthropogenic vegetation that occurs in small, highly fragmented, and often ephemeral stands influenced by human activities; (iii) any other vegetation types dependent on accidental disturbances and forming dynamic stages of secondary succession; (iv) vegetation of rock fissures and ledges, occurring on vertical surfaces, which are not depicted on maps. For such vegetation types, it is necessary to find a widely accepted way of quick *in situ* random selection of sampling plot site. Also, rapid progress of remote sensing techniques is promising in this respect. For example, high-resolution hyperspectral airborne imagery is able be provide stratification even at a very fine scale (SCHMIDTLEIN & SASSIN 2004).

(2) As follows from the above example, even well-prepared stratified random sampling plans can fail to record rare vegetation types, which are usually confined to specific habitats. It is therefore to be considered whether and to what extent stratified random sampling plans should be complemented with additional preferential sampling.

(3) Stratified random sampling plans select sites to be sampled irrespectively of their accessibility. However, in some wilderness areas, especially in the mountains, it may be too costly or even dangerous to sample vegetation in remote sites. Therefore in some cases sampling plans have to find a compromise between site accessibility and sampling representativeness (AUSTIN & HEYLIGERS 1989, WESSELS et al. 1998).

(4) High-quality environmental data necessary for habitat stratification are readily available just for a limited portion of the earth's surface, though the progress is incredibly rapid in this respect.

STATISTICAL ANALYSIS OF NON-RANDOMLY SAMPLED ECOLOGICAL DATA

The use of non-randomly sampled or spatially autocorrelated data is nearly inevitable in ecology. It would be unwise to dismiss such data just for the sake of methodological purism, because they can provide important insights into ecological patterns and processes. It is not a problem that these data are used and analyzed – generally any data may be useful if appropriately analyzed and interpreted. The problem is, however, that such data are routinely subjected to statistical tests which require independence of observations (LÁJER 2007).

Large-scale ecological phenomena are usually beyond the reach of controlled experiments. Therefore, many large-scale ecological analyses are of exploratory nature. Even if they start with an explicit hypothesis, it is extremely difficult to realize a sampling design in which only variables of interest are varied while others are held constant. The existence of complex interactions among environmental variables which cannot be manipulated on large scales often precludes the application of the traditional Popperian methodology based on hypothesis testing (POPPER 1990). Fortunately, methods of exploratory analysis or data mining developed in recent decades are able to reveal patterns in the data either without applying classical statistical tests or with the use of randomization procedures that distinguish real patterns from null expectations. Some methods of exploratory analysis are routinely used in ecology, particularly methods designed for simplification of complex multivariate relationships (multivariate classifications and ordinations; LEGENDRE & LEGENDRE 1998), while some others are still underappreciated by ecologists. In particular, the whole approach of algorithmic modelling as an alternative to stochastic modelling (BREIMAN 2001) offers powerful tools for the analysis of large-scale ecological data, including classification and regression trees (BREIMAN et al. 1984), artificial neural networks (RIPLEY 1996) or support vector machines (VAPNIK 1998).

CONCLUSIONS

We assert that to date we do not have an adequate substitute for the existing large databases of preferentially sampled phytosociological data for the use in large-scale vegetation studies (compare MARGULES & AUSTIN 1994 for the case of species distribution data). If we adhere to the requirements of methodological purists and try to make vegetation science "more exact" by abandonment of historical phytosociological data, we will restrict our ability to answer interesting ecological questions.

Ecologists should distinguish between the study of patterns and processes at small scale, where it is possible to find enough replicates, to use theoretically flawless sampling designs, to conduct exact measurements and to apply manipulative experiments, and the study of large-scale phenomena, which is more difficult to comprehend methodologically. Certain level of simplification and deviation from statistical dogmas is necessary at the large scales in many cases. Unless this distinction is appreciated, the study of large-scale ecological phenomena may be felt as doubtful or uninteresting by the ecologists trained in a spirit of

statistical purism. Taking into account the richness of ideas and the amount and quality of data gathered by phytosociology and other approaches to large-scale vegetation study during the last century, such development in the field of ecology would be alarming.

REFERENCES

AUSTIN M.P. & HEYLIGERS P.C. (1989): Vegetation survey design for conservation: gradsect sampling of forests in Northeast New South Wales. *Biol. Conservation* 50: 13–32.

BLACKBURN T.M. & GASTON K.J. (2003): *Macroecology. Concepts and consequences*. Cambridge University Press, Cambridge.

BOTTA-DUKÁT Z., CHYTRÝ M., HÁJKOVÁ P. & HAVLOVÁ M. (2005): Vegetation of lowland wet meadows along a climatic continentality gradient in Central Europe. *Preslia* 77: 89–111.

BRAUN-BLANQUET J. (1964): *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Ed. 3. Springer, Wien.

BREIMAN L. (2001): Statistical modelling: The two cultures. *Statist. Sci.* 16: 199–231.

BREIMAN L., FRIEDMAN J.H., OLSHEN R.A. & STONE C.G. (1984): *Classification and regression trees*. Wadsworth International Group, Belmont.

BROWN J. H. (1999): Macroecology: progress and prospect. *Oikos* 87: 3–14.

BUNCE R.G.H., BARR C.J., CLARKE R.T., HOWARD D.C. & LANE A.M.J. (1996): Land classification for strategic ecological survey. *J. Environm. Managem.* 47: 37–60.

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

CHYTRÝ M., PYŠEK P., TICHÝ L., KNOLLOVÁ I. & DANIHELKA J. (2005): Invasions by alien plants in the Czech Republic: a quantitative assessment across habitats. *Preslia* 77: 339–354.

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

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.

COCHRAN W.G. (1977): *Sampling techniques*. Ed. 3. Wiley, New York.

EWALD J. (2003): A critique for phytosociology. *J. Veg. Sci.* 14: 291–296.

FANELLI G., BIANCO M.P., CAZZAGON P., D'ANGELI D., DE CORSO S., DE SANCTIS M., GIOIA P., RAMELLO A., RINIERI G., SERAFINI SAULI A., TESCAROLLO P., TESTI A. & PIGNATTI S. (2005): Remote sensing in phytosociology: The map of vegetation of the Provincia of Rome. *Ann. Bot. (Rome), Nuova Ser.* 5: 171–181.

GOEDICKEMEIER I., WILDI O. & KIENAST F. (1997): Sampling for vegetation survey: some properties of a GIS-based stratification compared to other statistical sampling methods. *Coenoses* 12: 43–50.

GRABHERR G., REITER K. & WILLNER W. (2003): Towards objectivity in vegetation classification: the example of the Austrian forests. *Pl. Ecol.* 169: 21–34.

HAILA Y. & MARGULES C.R. (1996): Survey research in conservation biology. *Ecography* 19: 323–331.

HÁJEK M. & HÁJKOVÁ P. (2004): Environmental determinants of variation in Czech *Calthion* wet meadows: a synthesis of phytosociological data. *Phytocoenologia* 34: 33–54.

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

HIRZEL A. & GUISAN A. (2002): Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Modelling* 157: 331–341.

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

KENKEL N.C., JUHÁSZ-NAGY P. & PODANI J. (1989): On sampling procedures in population and community ecology. *Vegetatio* 83: 195–207.

KENT M. & COKER P. (1992): *Vegetation description and analysis. A practical approach*. CRC Press, Boca Raton.

KNOLLOVÁ I., CHYTRÝ M., TICHÝ L. & HÁJEK O. (2005): Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *J. Veg. Sci.* 16: 479–486.

KOENIG W.D. (1999): Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.* 14: 22–26.

LÁJER K. (2007): Statistical tests as inappropriate tools for data analysis performed on non-random samples of plant communities. *Folia Geobot.* 42: 115–122.

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. & MYERS D. (2002): The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25: 601–615.

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

LOSOSOVÁ Z., CHYTRÝ M., CIMALOVÁ Š., OTÝPKOVÁ Z., PYŠEK P. & TICHÝ L. (2006): Classification of weed vegetation of arable land in the Czech Republic and Slovakia. *Folia Geobot.* 41: 259–273.

LVONČÍK S. (2006): *Lesní vegetace Slezské pahorkatiny a Ostravské pánve* (*Forest vegetation of the Slezká Uplands and Ostrava Basin*). Diploma thesis, Masaryk University, Brno.

MARGULES C.R. & AUSTIN M.P. (1994): Biological models for monitoring species decline: the construction and use of data bases. *Philos. Trans. R. Soc. London B* 344: 69–75.

MOHLER C.L. (1983): Effect of sampling pattern on estimation of species distributions along gradients. *Vegetation* 54: 97–102.

MOORE J.J., FITZSIMONS P., LAMBE E. & WHITE J. (1970): A comparison and evaluation of some phytosociological techniques. *Vegetatio* 20: 1–20.

ØKLAND R.H. (2007): Wise use of statistical tools in ecological field studies. *Folia Geobot.* 42: 123–140.

PODANI J. (1984): Spatial processes in the analysis of vegetation. Theory and review. *Acta Bot. Hung.* 30: 75–118.

POPPER K.R. (1990): *The logic of scientific discovery*. Ed. 14. Unwin Hyman, London.

PYŠEK P., JAROŠÍK V., CHYTRÝ M., KROPÁČ Z., TICHÝ L. & WILD J. (2005): Alien plants in temperate weed communities: prehistoric and recent invaders occupy different habitats. *Ecology* 86: 772–785.

RICOTTA C. (2007): Random sampling does not exclude spatial dependence: the importance of neutral models for ecological hypothesis testing. *Folia Geobot.* 42: 153–160.

RIPLEY B.D. (1996): *Pattern recognition and neural networks*. Cambridge University Press, Cambridge.

SCHMIDTLEIN S. & SASSIN J. (2004): Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing Environm.* 92: 126–138.

SMART S.M., CLARKE R.T., VAN DE POLL H.M., ROBERTSON E.J., SHIELD E.R., BUNCE R.G.H. & MASKELL L.C. (2003): National-scale vegetation change across Britain; an analysis of sample-based surveillance data from the Countryside Surveys of 1990 and 1998. *J. Environm. Managem.* 67: 239–254.

STOHLGREN T.J., COUGHENOUR M.B., CHONG G.W., BINKLEY D., KALKHAN M.A., SCHELL L.D., BUCKLEY D.J. & BERRY J.K. (1997): Landscape analysis of plant diversity. *Landscape Ecol.* 12: 155–170.

TER BRAAK C.J.F. & ŠMILAUER P. (2002): *CANOCO reference manual and CanoDraw for Windows user's guide. Software for Canonical Community Ordination (version 4.5)*. Biometris, Wageningen & České Budějovice.

VAPNIK V. (1998): *Statistical learning theory*. Wiley, New York.

WESSELS K.J., VAN JAARSVELD A.S., GRIMBEEK J.D. & VAN DER LINDE M.J. (1998): An evaluation of the gradsect biological survey method. *Biodivers. & Conservation* 7: 1093–1121.

WESTHOFF V. & VAN DER MAAREL E. (1978): The Braun-Blanquet approach. In: WHITTAKER R.H. (ed.), *Classification of plant communities*, W. Junk, The Hague, pp. 287–399.