Subjectively Sampled Vegetation Data: Don't Throw Out the Baby With the Bath Water

Jan Lepš[1] **& Petr Šmilauer**[2]

1) Department of Botany, Faculty of Biological Sciences, University of South Bohemia, and Biological Centre, Institute of Entomology, CAS, Branišovská 31, CZ-370 05 České Budějovice, Czech Republic; e-mail suspa@bf.jcu.cz
2) Department of Ecology, Faculty of Biological Sciences, University of South Bohemia, Branišovská 31, CZ-370 05 České Budějovice, Czech Republic; e-mail petrsm@jcu.cz

Abstract: LÁJER (2007) reviewed the use of statistical methods for the analysis of vegetation data and concluded that their assumptions are often violated. Most importantly, the vegetation data are usually not a random sample of vegetation, and consequently the statistical tests should not be used at all. Although we appreciate many concerns raised in Lájer's paper, we find his conclusions far-fetched and inappropriate for practical use. In large-scale observational studies, it is practically impossible to define the sampled statistical population in the way that would allow for truly random sampling. As we are unable to ensure that any part of the population can appear in our sample with the same probability, we must try to avoid interaction between any bias imposed by the sampling procedure with the hypotheses we test. The same stance can also be taken in the case of datasets extracted from the databases of phytosociological relevés, where a substantial bias was introduced by phytosociologists' prejudice in selection of sampling locations. We can hardly use this kind of data when testing hypotheses concerning absolute values of species richness or discontinuities in community variation, but there are also many important research hypotheses that can be tested even with the help of this data source. In fact, such voluminous databases are the only available source of data for exploring research questions referring to large spatial or temporal scales. We must also reject Lájer's suggestion that application of descriptive, exploratory statistical methods is less harmful for datasets with sampling bias than application of tests of statistical hypotheses. Finally, we stress the difference between observational and experimental studies, because the rigor of sampling design should be respected much more in the case of latter ones.

Keywords: Assumption violation, Braun-Blanquet relevés, Experiments, Test robustness, Vegetation database

INTRODUCTION

The importance of data analysis and statistical inference has increased tremendously in most scientific fields in the past decades. However, application of statistics is based on a series of assumptions that are often not fulfilled in real data. LÁJER (2007) in his paper pointed out to some drawbacks in the use of statistical tests, particularly in vegetation science. His most important conclusion is that the application of any statistical test to data collected using the classical Braun-Blanquet method (where the records of vegetation composition are often called relevés), should be avoided because the data do not represent a random sample. According to Lájer, this is particularly true when analyzing relevés taken from vegetation databases. Instead, he recommends using various exploratory analysis techniques, and to

calculate indicator indices to ecologically interpret the data. If we would follow his recommendation, the thousands of relevés in vegetation databases should not be used beyond the simple exploration and calculation of environmental averages using the species indicator values (which, however, should not be analyzed any further, because they also do not fulfil the assumption of a random sampling).

LÁJER (2007) points out two assumptions that are usually not fulfilled in the analyses of phytosociological data – the classical relevés are not taken at randomly selected plots (and so the resulting set of relevés is usually not a random sample of the studied vegetation) and the data do not have desired distributional properties (normality and homoscedasticity). The latter assumption (distributional properties of ecological data) was discussed several times in the literature: an ecologist dealing with this issue should read the Special Feature and Forum in the journal *Ecology* (e.g. POTVIN & ROFF 1993, JOHNSON 1995, SMITH 1995, STEWART-OATEN 1995). In addition to the obvious fact that the opinion about normality violation differs among scientists, the reader will learn how the robustness of the statistical tests depends on sample size – the larger a sample is, the more robust is the method against deviation from normality. In addition, since 1995 many new methods have been developed or have become more widely used (namely regression methods relying on a more general assumption of distributions from an exponential family, such as the Generalized Linear Models, Generalized Additive Models, or Generalized Mixed Effect Models) and these methods are a natural choice for many types of data with non-normal distributions or with the variance changing with the expected response value. Also, requirements concerning the type of distribution are much ameliorated in permutation tests, despite such tests sharing some sensitivity to outliers or heterogeneity of variation (ANDERSON & TER BRAAK 2003). Consequently, we focus here on the problem of non-random sampling, which we see as more serious.

We try to demonstrate that the habit of applying statistical tests to data that are not a real random sample is quite common in various biological disciplines under circumstances where collecting a real random sample is difficult or not feasible at all. Also by the analogy with these applications, we claim that the sampling that is neutral with respect to the tested hypotheses leads to acceptable datasets for statistical testing. We also suggest that there is no difference between exploratory statistical methods and methods of statistical inference in whether subjectively sampled vegetation data can validly be used. On the other hand, greater rigour is needed in experimental, manipulative studies.

HOW TO GET A RANDOM SAMPLE

In Europe, classical phytosociological relevés are most often used to describe and analyze patterns on a broad spatial scale. A typical example might be a study of the dependence of species composition (a multivariate response) or some diversity measure (a univariate response) on measured environmental factors in the meadows of Central Europe. The whole task is seemingly simple – we just randomly select the locations, record the relevés, measure environmental characteristics there, and finally we test the relationships with appropriate statistical methods. So, why are studies following this approach so rare (at least at the large spatial scales)?

The real problem is to get a random sample. To obtain a truly random sample, we must first define precisely the statistical population from which our sample should be taken and to which our conclusions will apply. In practice this means that we need to define exact geographical boundaries of the area and then an operational definition what should be considered a meadow in our study. Then, we can continue in two ways. We can send several teams of ecologists to prepare maps showing all the meadows within the delimited area (or use the information obtained by remote-sensing method and stored in a GIS, see e.g. GRABHERR et al. 2003), and then use some random process, generating coordinates of plots to be sampled. Alternatively, we can first generate the coordinates of all the required plots, then visit suggested locations and decide on the spot, whether the plot complies to our definition of "meadow", and record a relevé at each of the "accepted" plots. This procedure should be of course repeated, until the desired number of relevés is obtained. In most cases, such approaches are difficult to apply.

So what research possibilities remain? In the European context, extensive sampling campaigns with proper design and great geographical extent are lacking, with just a few notable exceptions (e.g GRABHERR et al. 2003). Therefore, we must either obtain the data from an existing vegetation database and proceed with caution, taking into account the dangers issuing from the way the data were collected (particularly the subjective selection of sampled locations), or we abandon statistical analyses at this spatial scale and possibly decide to work on a spatially limited scale, where random sampling is feasible. We consider the first option to be a viable approach and unlike LÁJER (2007), we see the use of statistical methods under such circumstances as a consequence of realizing that at the large scales we do not have any better choice.

THEORETICAL ASSUMPTIONS AND THE REALITY

All statistical tests are based on some assumptions. If the assumptions are met, the results of statistical tests are valid. However, if the assumptions are not fulfilled, the results may, but need not be biased. The degree to which the test results are resistant to violation of their assumptions is called test robustness. From a practical point of view, it is one of the most important properties of statistical tests applied to biological data collected in field, because the assumptions are only rarely fulfilled completely.

In our view, LÁJER (2007) ignores the issue of the extent of bias introduced by using the data collected by a non-random selection, despite promising to the reader in the Introduction section that "it is examined to which extent the results obtained... are reliable". Instead, he takes a stance that can be paraphrased as "if the assumptions are not completely met, the results are completely wrong". LÁJER (2007) also assumes that only a random sample, and not a sample taken according to some pattern is correct for the use of the chi-square test. One of the assumptions of the chi-square test is that the sampling units are selected randomly from an infinite population. Although not explicitly stated, all the mathematical derivations in the appendices of Lájer paper are based on the assumption of sampling an infinite basic population. If we strictly follow Lájer's logic (i.e., when the assumptions are not fulfilled, a test is not valid), the chi-square test should not be applied in any field study, because the assumption of infinity is never fulfilled. However, we (similarly to other researchers) believe

that reasonably large sampled populations are a sufficient approximation to infinity. Nevertheless, if we really had an infinitely large area (which is the only possibility to obtain an infinitely large population of non-overlapping sampling units), the probability that two selected sampling units will be close to each other, and thus the results will be affected by their spatial dependence, should be practically equal to zero. So if we work with meadow vegetation of infinite size, we do not need to care about two randomly selected plots falling into the same patch of, say, invading *Calamagrostis epigeios*. But in our real sized meadows, we have to. In the real world, use of the sampling schemes preventing plots to appear close to each other is a reasonable concern. In fact, an unbiased sampling scheme, other than a completely random placement of plots, might be preferable, depending on the extent and form of spatial correlation (LEGENDRE et al. 2005).

Unlike LÁJER (2007), we see real sampling as a process, where all the theoretical assumptions are usually not fulfilled completely, but where we simultaneously try to minimize the possible dangers issuing from the violation of various assumptions. We believe that the general role of statistical testing is to guard against over-interpretation of observed patterns. If the test shows that a relation is not statistically significant, we abstain from its interpretation. If the test is significant, we can start thinking about interpreting our findings. Of course, such interpretation should include any possible bias and any confounding effects as well as that any positive correlation among the sampling units makes our test too liberal. Such an approach to statistical testing is particularly appropriate when the number of sampling units is sufficiently large, and so the power of our tests is reasonably high. This approach also makes it clear that a statistical test is never the end of the story. If it is, the story is probably uninteresting anyway.

VIOLATION OF RANDOMNESS AND ITS CONSEQUENCES: NOT JUST IN VEGETATION SCIENCE

The assumption of randomness and independence of sampling units in the sample is one of the principal assumptions of statistical inference. But in many scientific fields (not only the phytosociology) obtaining a real random sample is difficult, if not impossible. In the excellent paper of STEWART-OATEN (1995), we can read: " 'Sensible' opinions are treated as mandatory rules, frustrating authors... One possible reason for this is an aura of exactitude and universality inherited from mathematics, although all interference is approximate not only because of the use of limiting distributions but also because of the model uncertainty, treatment-subject non-additivity, *and sampling that is random in at best a limited slice of time and space*." (emphasis by JL & PŠ). So, in our view, it is better to accept the fact that our sampling is not really random and interpret our conclusions as approximate ones instead of dividing our problems into those where we pretend that the sampling is completely random, and those where no statistical inference is possible. This does not mean that we should not carry out a random sampling where it is feasible and do the best to obtain a sample as close to random as possible in the other situations. Just to follow another recommendation of STEWART-OATEN (1995): "Do the best you can."

Random sampling ensures that each sampling unit has the same probability of being selected into the sample and that a selection of one sampling unit does not affect the probability that another unit will be selected (assumption of the independence of units).

Accordingly, there are generally two ways in which this assumption can be violated. First, sampling can be biased, e.g. when a phytosociologist selects plots that are more species rich, or that are "typical" ones, based on his *a priori* views – this means that some sampling units (e.g. those considered typical) have higher probability of being included into the sample. Secondly, there might be a dependency between some of the plots, e.g. when the collector reaches a locality, he usually takes more than one relevé. This means that the individual sampling units are not independent – selecting a plot increases the probability that another one (or more) will be recorded close to the first plot. Such an effect is even more pronounced in phytosociological databases: it is well known that the existing landscape is not covered evenly by the existing records – samples are clumped in particular areas which are either close to the homes of active researchers or in otherwise attractive locations. This difficulty can be partially circumvented by a proper selection of records from the databases (e.g. using geographical stratification; KNOLLOVÁ et al. 2005).

The problem of obtaining a real random sample is not restricted to vegetation science – it is common in other fields of ecology (and science in general). The examples below are not intended as justification of the type "When others do it wrong, why should not we be allowed to do it wrong as well?", but just as an example that other biological fields face similar problems, and generally nowhere would researchers decide to refrain from statistical testing. Instead, the statistical tests are applied, and the possible consequences are taken into account in the interpretation of results.

In some fields, truly random sampling is even more difficult than for the vegetation data[1]. Imagine an entomological study, where we need a random sample of individuals of a moth population in a patch of forest. To obtain it, we would first need to collect all the individuals of a given population, then assign each of them a number from 1 to N, where N is the number of all individuals in the population, and only then decide, using a random number generator, which to include in our random sample. This is clearly not feasible. Instead, we must install a light trap and consider the individuals that were caught to be a random sample. However, one should be aware that a bias might be introduced – various properties of an individual affect its ability to reach the light trap. Consequently, we should never use this sample for estimating anything connected with the insect sensitivity to light (if there is a fraction of individuals not attracted to light, it is naturally not in the sample). We should also be extremely cautious of estimating the size distribution of individuals in this population, because we would be afraid that the larger individuals are better fliers, and consequently are attracted from larger distances than the smaller individuals. However, to judge this danger, we would need to know a lot about the biology of the species (e.g., is the ability to fly correlated with the size of the individual?). Finally, we would not hesitate to use this sample as a random sample for

---

1 We would even say that, in a sense, the vegetation scientists are generally in a better position than others, because the fact that relevé collection is not completely random is generally known and often taken into account in interpretations. This is not the case in some other fields. Just try to remember how often population ecology studies claim that the measurements were done on "randomly selected individuals". We doubt that in all these cases, the statistical population was really first defined (if not, random selection of individuals is impossible) and that the correct method ensuring for each individual an equal and independent probability to be selected was applied.

estimating the relationship between biomass of a female and number of eggs in its ovaries (and probably, the validity of a test of the null hypothesis of no correlation would be hardly questioned by anybody).We would do so simply because we (and hopefully the reviewers and readers of our paper) believe that the sampling method was not biased with respect to the relationship measured. Whereas in a random sample we would have statistically guaranteed that the bivariate distribution of weight and number of eggs is an unbiased estimate of the distribution in the whole population, the statistical inference based on our real sample is based on the assumption that the sampling procedure is neutral to this relationship.

Hence, when we speak about a bias in sampling procedure, such a bias must be evaluated with respect to the tested statistical hypotheses. Our "sampling neutrality" relates to the idea of exchangeability, which is the fundamental requirement for sampling units to be appropriate for statistical analysis. Randomized experiments (including randomization of sampling positions), if possible, provide unconditional exchangeability for our data. Nevertheless, we can also make a reasonable statistical inference using a dataset that is conditionally exchangeable (GELMAN et al. 1995), if we can express the sampling design (such as the spatial arrangement of our plots in a field) as a set of covariates included in our model.

As another example, it often happens in experimental field ecology that the same experiment, if executed at different sites, gives contradictory results. Therefore, one should be very careful when interpreting results obtained from a single site and so the multi-site experiments receive more and more attention (HECTOR et al. 1999, VAN DER PUTTEN et al. 2000). Let us consider, for simplicity, an experiment involving a single experimental factor, e.g. with several levels or types of fertilization, and with the fertilizing effect evaluated using a univariate response variable. If performed at a single site, arranging for a correct experimental design is not difficult. However, test results apply just to this single experimental site, and any generalization beyond it is then a non-statistical reasoning based on our belief that what is true in our experimental site, is also true for the other sites with comparable conditions. To cover a broader selection of habitat conditions, we might prefer to repeat such a simple experiment at multiple locations, analyzing the data with a two-way analysis of variance with the treatment being a fixed factor and the site identity a random factor (e.g. VAN DER PUTTEN et al. 2000). In this case, the results can be generalized to the habitats represented by our experimental sites. But for a random factor, its levels are considered to be a random selection of all possible ones, i.e., our experimental sites are considered to be a random selection from all the possible sites of that habitat type. This means that to fulfill the condition of random selection, the experimental sites should be selected by an appropriate random process – basically in a way described in the "How to get a random sample" section[2].

The larger the temporal and spatial scale we approach, the more difficult it is to design a sampling scheme that would be both statistically correct and logistically feasible at the same time. The same is true for any progressive generalization. This can be nicely seen in the use of meta-analyses in ecology (GUREVITCH & HEDGES 2001). Ecological experiments are not done at randomly selected experimental sites, and moreover, their publication is highly

---

2   Just as a corollary, land owner's unwillingness to rent the site we randomly selected for our experimental purposes would introduce an important bias – the experimental use of more productive sites would be probably rejected more often.

selective (toward those showing results "interesting" to individual journal readers). Still, meta-analyses are carried out, because they are probably the best possibility to synthesize the results of independent experiments. Also, possible consequences of bias in selection of publications and experimental sites are taken into account when interpreting meta-analyses results.

BROAD-SCALE PATTERNS IN VEGETATION SCIENCE

Considerations outlined in the previous examples apply similarly (and mostly implicitly) to the data stored in phytosociological databases. Let us imagine that we want to address some research hypothesis in the context of seminatural meadows, classified as the *Molinio-Arrhenatheretea* class in Central Europe. As it was shown above, to obtain a real random sample is difficult. As an alternative, we can obtain a dataset from a phytosociological database, but the sample obtained in this way is clearly not random. Nevertheless, behaviour of a phytosociologist (as a sort of scientist) during the collection of data that became part of such a database is at least partially predictable. Therefore, we shall not try to estimate average species richness within this vegetation type, because our phytosociologist intentionally disregards very species-poor locations and possibly also very rich ones that fall into areas of high environmental heterogeneity that he would see as a "transition" between *a priori* conceived types. Also, based on the knowledge of a phytosociologist's behaviour, we should expect that the strength of interspecific associations is exaggerated: a phytosociologist probably selects "typical" vegetation types (where species combinations, which are "typical" based on his or her previous experience, are seen). This behaviour leads to the bias mentioned by LÁJER (2007).

Interestingly, the idea of incorporating sampling bias (by modelling "the data collection mechanism") in the analyses is a natural part of the Bayesian approach to statistical analysis (GELMAN et al. 1995, chapter 7).

With all the hesitation outlined above, we still agree with LÁJER (2007) that there are many research questions where using the data collected with the aim of classifying vegetation types is clearly not appropriate, despite attempts in several published studies. But many opportunities where our hypothetical dataset of phytosociological relevés obtained from a database could be used for testing hypotheses still exist. For example, if we have measurements of soil characteristics available for some relevés, we can reasonably believe that the phytosociologists' choice of plots was neutral (unbiased) with respect to the relationship between soil characteristics and species composition, and consequently, with some caution, we should be able to test the significance of this relationship. It is, however, quite natural that random and phytosociological sampling will find different dominant compositional gradients (revealed, e.g. by the first DCA axis of the data set), even when sampling the same landscape, and those different gradients will naturally correlate with different environmental variables (as in DIEKMANN et al. 2007). In addition, there are also many interesting research questions that encompass a very broad spatial and/or temporal scale, where large databases of phytosociological data are the only reasonable resource and where the relevés can be used in a valid way, because the present sampling bias does not interact with the tested hypothesis.

With the increasing spatial extent of our sample, we can still try to prepare a sampling plan approaching random (or other unbiased) sampling, but the problem of clearly delineating the scope for our conclusions (i.e., defining the sampled population) becomes in practice more and more difficult (and costly). Whereas there were no problems with designing a probabilistic sampling plan in a rather homogeneous 113 ha continuous ancient forest (BORCHSENIUS et al. 2004), sampling the entire forest area of Austria (GRABHERR et al. 2003) was undoubtedly logistically more difficult. This study demonstrates the possibility to carry out a probabilistic sampling program, particularly by taking advantage of modern GIS technologies. However, even in this case some locations had to be rejected by rather subjective criterion (tree layer strongly altered by forestry). Nevertheless, such approaches are very promising, and, very probably, some data sets collected in a statistically correct manner from forest inventories already exist, particularly in areas with less fragmented forests than in Central Europe. One should, however, realize that similar approaches would be very difficult to apply in studies of vegetation types of highly fragmented distribution, the extent of which could hardly be precisely determined from available GIS databases or remote sensing data (like, e.g., the *Sphagnum*-rich mires in HÁJKOVÁ & HÁJEK 2004).

We will also face a decision concerning the trade-off between sample size and scope versus its quality. We should consider the question whether it is worth designing an expensive sampling scheme, which would be necessarily restricted in spatial extent and the number of collected relevés, in comparison with using what is already available in vegetation databases. Facing this dilemma, we would suggest analyzing what is available in the databases in an exploratory manner and then, based on such results, design a specific sampling plan as close to random sampling as possible, focused on testing the hypotheses suggested by the preceding exploratory analysis (following partially the logic of HALLGREN et al. 1999).

ARE THE EXPLORATORY METHODS REALLY LESS AFFECTED BY NON-RANDOM SAMPLING THAN TESTS ARE?

We must also question the recommendation of LÁJER (2007) concerning data collected with non-random sampling, in which he suggests abstaining from statistical testing and using only the techniques of exploratory data analysis. We demonstrate our concerns by focusing on Lájer's recommendation to use only the unconstrained ordination, and not the constrained ordination for such sets of relevés. First we note that the choice of constrained ordination does not force its user to perform a statistical test – constrained ordination can be used without subsequent testing, even though many of its users focus primarily on the results of a significance test, estimating the probability of committing an error by rejecting a true null hypothesis. But from a practical view, we should realize that the test of significance might not be the most important result of the constrained ordination: a null hypothesis about a complete independence of community composition on the widely collected environmental characteristics is patently unrealistic and its rejection is just a matter of the test power. If we would believe, however, that our sampling is so biased that we cannot use any statistical tests, then we do not see any reason to believe that an informal interpretation of an unconstrained ordination diagram would also not be biased. In particular, if the relevé locations were selected as "typical" plots, then the unconstrained ordination diagram will show clear discontinuities (being just an artifact of sampling method, not reality), whereas the

relationship between species composition and environment (e.g. relationships of species to environmental gradients) might be fairly less affected. A similar concern arises for other exploratory methods – if our selection of plots is biased toward the species-rich ones, there is no reason to expect that the stem and leaf display will be less biased than any test we can apply to species-richness estimates. In our view, no general rule exists here – the danger arising from the sampling method should always be taken into account in our interpretations, and evaluated within the context of the questions we ask.

OBSERVATIONAL AND EXPERIMENTAL STUDIES

The "liberal" approach to assumption violation that we advocate here is aimed mainly at observational studies, particularly those carried out on large spatial scales. We should be much more strict in the case of experimental studies, particularly in the case of small-scale manipulative experiments, where the correct experimental design is applicable, and that were designed to test specific, *a priori* stated hypotheses. In observational studies, causal interpretation of their results must be always very careful and, within the framework of scientific reasoning, such studies should be considered mainly as a tool offering mechanistic explanations that should be further subjected to experimental confirmation. Causal interpretation of manipulative experiments, however, is much stronger (GOTTELI & ELLISON 2004), and we should be therefore more strict in their design so as not to compromise the strength and validity of our inference. However, even in manipulative experiments we are restricted by the feasibility of experimental manipulations, determined mostly by their spatio-temporal scale (GOTTELI & ELLISON 2004). Twenty years ago DIAMOND (1986) demonstrated the trade-off between the control of independent variables and the correctness of experimental design on one hand, and the achievable spatio-temporal scale and realism on the other hand. He also advocated that combinations of "strict" methods be applied on smaller scales with other studies (like "natural experiments") that do not have a fully correct statistical design, but are more realistic and cover a wide range of spatial and temporal scales.

CONCLUSIONS

We are convinced that proper random sampling is logistically hardly achievable for some important questions about vegetation patterns. If we accept this postulate, we have three possibilities how to proceed: (1) stop asking such questions, (2) ask these questions, but abstain from statistical analyses, (3) use the data that can be collected in a feasible way, even if they do not conform to the strict statistical requirements, and take into account all the dangers in interpretation of our results. Of the three options, we clearly prefer the third one.

Recent vegetation databases contain tens (and probably hundreds) of thousands of relevés, collected over nearly a century. Despite all the drawbacks issuing from the manner in which the data were collected, we believe that they are source of extremely valuable information that can and should be explored with a range of statistical methods. The fact that they are not a result of random sampling should be, however, taken into account very seriously when interpreting results. One should be aware that the extent of undesired impact depends on the hypotheses we test, and that there is no easy way to avoid this impact. As the only general rule,

we must repeat the STEWART-OATEN'S (1995) one: "Do the best you can". And with respect to Lájer's paper we add: do not throw out the baby with the bath water.

REFERENCES

ANDERSON M.J. & TER BRAAK C.J.F. (2003): Permutation tests for multi-factorial analysis of variance. *J. Statist. Comput. Simul.* 73: 85–113.

BORCHSENIUS F., NIELSEN P.K. & LAWESSON J.E. (2004): Vegetation structure and diversity of an ancient temperate deciduous forest in SW Denmark. *Pl. Ecol.* 175: 121–135.

DIAMOND J. (1986): Overview: laboratory experiments, field experiments, and natural experiments. In: DIAMOND J. & CASE TJ. (eds.), *Community ecology*, Harper & Row, New York, pp. 3–22.

DIEKMANN M., KÜHNE A. & ISERMANN M. (2007): Random vs non-random sampling: effects on patterns of species abundance, species richness and vegetation-environment relationships. *Folia Geobot.* 42: 179–190.

GELMAN A., CARLIN J.B., STERN H.S. & RUBIN D.B. (1995): *Bayesian data analysis*. Chapman & Hall, London.

GOTTELI N.J. & ELLISON A.M. (2004): *A primer of ecological statistics*. Sinauer, Sunderland.

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

GUREVITCH J. & HEDGES L.V. (2001): Meta-analysis. Combining the results of independent experiments. In: SCHEINER S.M. & GUREVITCH J. (eds.), *Design and analysis of ecological experiments*, Oxford University Press, pp. 347–369.

HÁJKOVÁ P. & HÁJEK M. (2004): Bryophyte and vascular plant responses to base-richness and water level gradients in Western Carpathian *Sphagnum*-rich mires. *Folia Geobot.* 39: 335–351.

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.

HECTOR A., SCHMID B., BEIERKUHNLEIN C., CALDEIRA M.C., DIEMER M., DIMITRAKOPOULOS P.G., FINN J.A., FREITAS H., GILLER P.S., GOOD J., HARRIS R., HOGBERG P., HUSS-DANELL K., JOSHI J., JUMPPONEN A., KORNER C., LEADLEY P.W., LOREAU M., MINNS A., MULDER C.P.H., O'DONOVAN G., OTWAY S.J., PEREIRA J.S., PRINZ A., READ D.J., SCHERER-LORENZEN M., SCHULZE E.D., SIAMANTZIOURAS A.S.D., SPEHN E.M., TERRY A.C., TROUMBIS A.Y., WOODWARD F.I., YACHI S. & LAWTON J.H. (1999): Plant diversity and productivity experiments in European grasslands. *Science* 286: 1123–1127.

JOHNSON D.H. (1995): Statistical sirens: the allure of nonparametrics. *Ecology* 76: 1998–2000.

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.

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., DALE M.R.T., FORTIN M.J., CASGRAIN P. & GUREVITCH J. (2005): Effects of spatial structures on the results of field experiments. *Ecology* 85: 3202–3214.

POTVIN C. & ROFF D.A. (1993): Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology* 74: 1617–1628.

SMITH S. M. (1995): Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology* 76: 1997–1998.

STEWART-OATEN A. (1995): Rules and judgments in statistics: three examples. *Ecology* 76: 2001–2009.

VAN DER PUTTEN W.H., MORTIMER S.R., HEDLUND K., VAN DIJK C., BROWN V.K., LEPŠ J., RODRIGUEZ-BARRUECO C., ROY J., LEN T.A.D., GORMSEN D., KORTHALS G.W., LAVOREL S., REGINA I.S. & ŠMILAUER P. (2000): Plant species diversity as a driver of early succession in abandoned fields: a multi-site approach. *Oecologia* 124: 91–99.