

Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species

Zuzana Münzbergová and Tomáš Herben

Münzbergová, Z. and Herben, T. 2004. Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. – *Oikos* 105: 408–414.

This study proposed a new quantitative technique to identify suitable but unoccupied habitats for metapopulation studies in plants. It is based on species composition at the habitat and knowledge of species co-occurrence patterns. It uses data from a large phytosociological database as a background for estimating species co-occurrence patterns. If such a database is not available, the technique can still be applied using the same data for which the prediction is done to estimate the species co-occurrence pattern. Using the technique we were able to indicate suitable unoccupied habitats and differentiate them from the unoccupied unsuitable ones. We also identified habitats with low probability of being suitable that were occupied. Compared to a direct approach of identification of suitable habitats, which involves introduction of a species to the habitat and studying its performance, the approach presented here is much easier to apply and can provide extensive information on habitat suitability for a range of species with much less effort and time needed.

Z. Münzbergová and T. Herben, Inst. of Botany, Academy of Sciences of the Czech Republic, CZ-252 43 Práhonice, Czech Republic (zuzmun@natur.cuni.cz) and Dept of Botany, Faculty of Science, Charles Univ., Benátská 2, CZ-12801 Praha 2, Czech Republic.

One of the key predictions of metapopulation theory is the occurrence of suitable, but unoccupied habitats (Levins 1969, Hanski and Simberloff 1997); due to the ongoing colonization and extinction, a certain fraction of suitable habitats should be unoccupied if the system is in equilibrium. Identification of unoccupied but suitable habitats is therefore a part of many empirical metapopulation studies. It is, however, often difficult to tell what constitutes a suitable habitat (Ehrlén and Eriksson 2000). This difficulty was recently used as one of the important arguments in the critique of application of metapopulation theory in plants (Freckleton and Watkinson 2002).

The most direct but technically difficult way for estimation is to introduce the species into the habitat and follow its performance (Eriksson and Ehrlén 1992,

Turnbull et al. 2000). To get a good estimate of suitability using this method one should follow the whole life cycle of the species as population bottlenecks may occur at later stages of recruitment (Losos 1995, Gustafsson et al. 2002). This can, however, take much longer than any research project can last (Ehrlén and Eriksson 2000).

Therefore, alternative indirect ways to estimate habitat suitability are sought (Husband and Barrett 1996). A common indirect way to define suitable habitats is to use environmental variables (Ouborg 1993, Husband and Barrett 1996, Quintana-Ascencio and Menges 1996). However, the array of environmental variables that can be measured is very wide (e.g. soil properties, microclimate, occurrence of extreme events) and there is no good way to decide which variables are the right ones.

Accepted 6 October 2003

Copyright © OIKOS 2004
ISSN 0030-1299

Measuring all of them is very laborious and time consuming so suitability estimation using this approach is of limited use (Ehrlén and Eriksson 2000).

An alternative approach is to use information on presence of other species as a measure of habitat suitability. It is based on the common experience that occurrences of species are correlated. Combination of a species list from a habitat and data on species co-occurrence then provides estimation of habitat suitability for the species of interest. Compared to the use of environmental variables this enables much better exploitation of information about conditions at a habitat (Ewald 2002) as species composition is an outcome of the important environmental variables. Reasoning of this type is the basis of phytosociology and is sometimes applied for smoothing of phytosociological databases (i.e. adding species that might have been overlooked) (Beals 1984) or gap filling of flora databases (Witte 1998). Recently a similar approach has been used to define species pool composition (Ewald 2002, Chytrý et al. in press). However, no attempt has been made to use it in a metapopulation context to define suitable unoccupied habitats.

The aim of this study is to explore patterns of habitat suitability for a particular species predicted from its co-occurrence with other species present at that habitat. This can be done by calculating the probability of species occurrence at a habitat regardless of its actual occurrence. If there are suitable but unoccupied habitats available, one would expect to get high probabilities of species occurrence at occupied habitats and both high and low probabilities at unoccupied habitats. Our prime question is therefore (i) is it possible to find the predicted pattern of estimated probabilities at occupied and unoccupied habitats (high probability at occupied and both high and low probability at unoccupied habitats) in a system of plant species on habitat islands?

Further, we are asking two methodological questions. First, a typical approach to calculate such probability is to use the same data-set for calibration and for prediction (Beals 1984, Witte 1998). However, circularity can be avoided if different data-sets are used to estimate species co-occurrence patterns and for predictions. In some countries there are large phytosociological databases that could be used as reference databases for this prediction (Ewald 2002, Chytrý and Rafajová 2003). When using a reference database there are many options of how large range of communities should be included in it. Therefore we ask (ii) how does the prediction of habitat suitability depend on the spectra of communities covered by the reference database?

Second, in some cases such a reference database may however not be available. Therefore, our last question is (iii) does the prediction change if the same database is used for both calibration and prediction?

To estimate habitat suitability we used Beals index of sociological favourability (Beals 1984), which defines probability of species occurrence based on co-occurrence of that species with other species. This index provides an estimate of the probability to encounter a species at a habitat. The estimate is independent of the fact whether the species really occurs there. To predict species occurrence we used data from dry grasslands in northern Bohemia, Czech Republic. Relevés from the Czech national phytosociological database were used as a reference database (Chytrý and Rafajová 2003).

Methods

Field data

To investigate the pattern of estimated probability at occupied and unoccupied habitats we used data from twenty-two localities of dry grasslands of the *Bromion* community. At each locality presence of all species was recorded in ten 1 × 1 m plots selected to cover the whole range of the locality. These data were then pooled and used as information on species composition of the locality.

Reference database

To estimate species co-occurrence patterns we used relevés from the Czech national phytosociological database (<http://www.sci.muni.cz/botany/database.htm>). We used relevés from the area delimited by longitude 13°15'–15°30' E and latitude 50°00'–50°40' N. To estimate the effect of reference database width on the prediction we used two different definitions of the reference database, a wider definition including all dry and mesophyllous grassland communities and a narrower definition including only broad leaf dry grasslands (*Bromion* community).

The wider database included all relevés of classes *Festuco-Brometea*, *Trifolio-Geranietea*, *Molinio-Arrhenatheretea*, *Artemisietea vulgaris*, *Agropyretea repens* and alliance *Agrostietalia*. It included 2784 relevés. The narrower database included only relevés of *Bromion* community in a narrower sense. The inclusion of relevés into this selection was based on presence of at least 5 diagnostic species of this community. The diagnostic species were selected as diagnostic species as defined by Chytrý (unpubl.) having the $\Phi > 0.30$ (Chytrý et al. 2002). It included 785 relevés.

To simulate a situation when reference database is not available we also used the calibrated set itself as the reference database. In this case the sample being calibrated was excluded from the data-set.

Probability of species occurrence

Probability of species occurrence at a habitat was estimated using Beals index of sociological favourability (Beals 1984). The probability of occurrence of species j at a habitat i is defined as

$$p_{ij} = (1/S_i) \sum_k N_{jk} / N_k \quad (1)$$

where p_{ij} is probability to find species j at habitat i , S_i is number of species at habitat i (minus 1 if species j is present), N_{jk} is number of joint occurrences of species j and k , $j \neq k$ in the reference database, N_k is number of occurrences of species k in the reference database. It is important to note that the Beals index value is independent of the actual occurrence of the species at a habitat. We calculated Beals index for each species in the data-set in two ways, using the external database and without it. We compared Beals index values for 45 species that were present in at least 3 localities.

Probability of occurrence at unoccupied habitats

Probability of occurrence at unoccupied habitats was estimated from a cumulative frequency distribution of Beals index values of each species at all occupied habitats. We used it to demonstrate where in this distribution each unoccupied habitat lies, i.e. where it would rank if it were occupied. This recalculation was done in order to define threshold value for suitability of the unoccupied habitats as the absolute values of the Beals index value strongly depend on the frequency of the species in the external database. These values (later called percentiles of Beals index value) can be directly interpreted in terms of suitability: an unoccupied habitat having Beals index value lower than any occupied habitat had a 0% probability of being occupied; a habitat having Beals index value equal to median of the Beals index values at the occupied habitats had 50% probability etc. (Fig. 1).

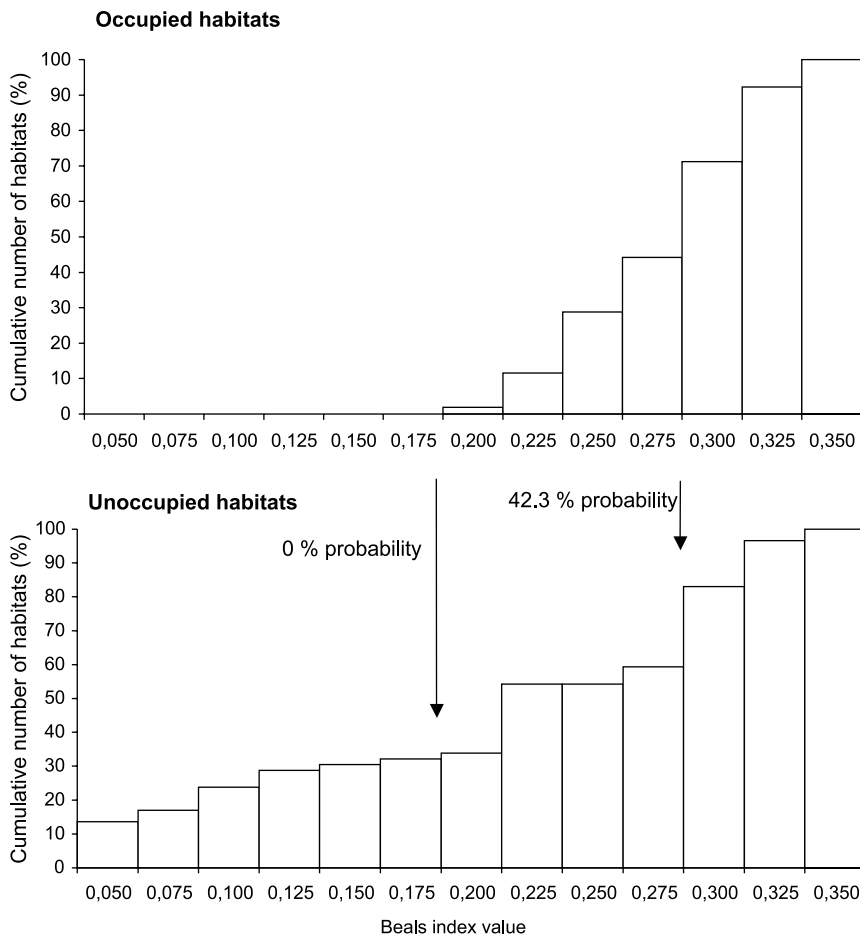


Fig. 1. An example of translation of Beals index values at unoccupied habitats into percentiles of Beals index values at occupied habitat. The upper histogram represents cumulative distribution of Beals index values at occupied habitats. The lower represents distribution of Beals index values at unoccupied habitats. The arrows indicate the probability that an unoccupied habitat with given probability will be occupied that was derived from the distribution of Beals index values at the occupied habitats. The scale used in actual calculations was much finer.

Reliability of estimation

To see how sensitive the Beals index value is to composition of the reference database we calculated 95% bootstrap confidence intervals of the value based on 100 resamplings of each reference database. Using these data we also calculated 95% confidence intervals of percentiles of Beals index value at unoccupied habitats.

Data analysis

We used linear regression to estimate the relationship between the three types of predictions (using two types of external database and without it). Beals index value and percentile of Beals index value calculated without external database were used as dependent variables (one data point corresponds to one species at one locality) and the respective value calculated using wider external database and species were used as independent variable. Comparison of the two external databases was done in the same way. The calculations were done using SPSS version 11.0.1.

Results

Estimation of probabilities

In all cases the Beals index value was on average higher at occupied than at unoccupied habitats (Fig. 2, Table 1), but there were some unoccupied habitats with percentile of Beals index value equal to 100 (the Beals index value was as high or higher than the highest Beals index value at any occupied habitat). This pattern was obtained from calculations using all three ways to estimate species co-occurrence patterns. In a few cases we also identified localities with low probability of being suitable that were actually occupied.

There is only a very weak but significant relationship between percentiles of Beals index values estimated using wider external reference database and using the calibrated set as reference database ($R^2 = 0.02$, $df = 1,675$, $F = 51.07$, $p = 0.024$). The relationship was however much stronger between the original Beals index values ($R^2 = 0.12$, $df = 1,989$, $F = 222.283$, $p < 0.001$). There was strong relationship between both the percentiles of Beals index value ($R^2 = 0.74$, $df = 1,675$, $F = 1351$, $p < 0.001$) and Beals index values ($R^2 = 0.99$, $df = 1,989$, $F = 7444$, $p < 0.001$) estimated using the two types of external reference databases.

Reliability of estimation of the probabilities

When the external reference database was not available the standard error values were larger than when external

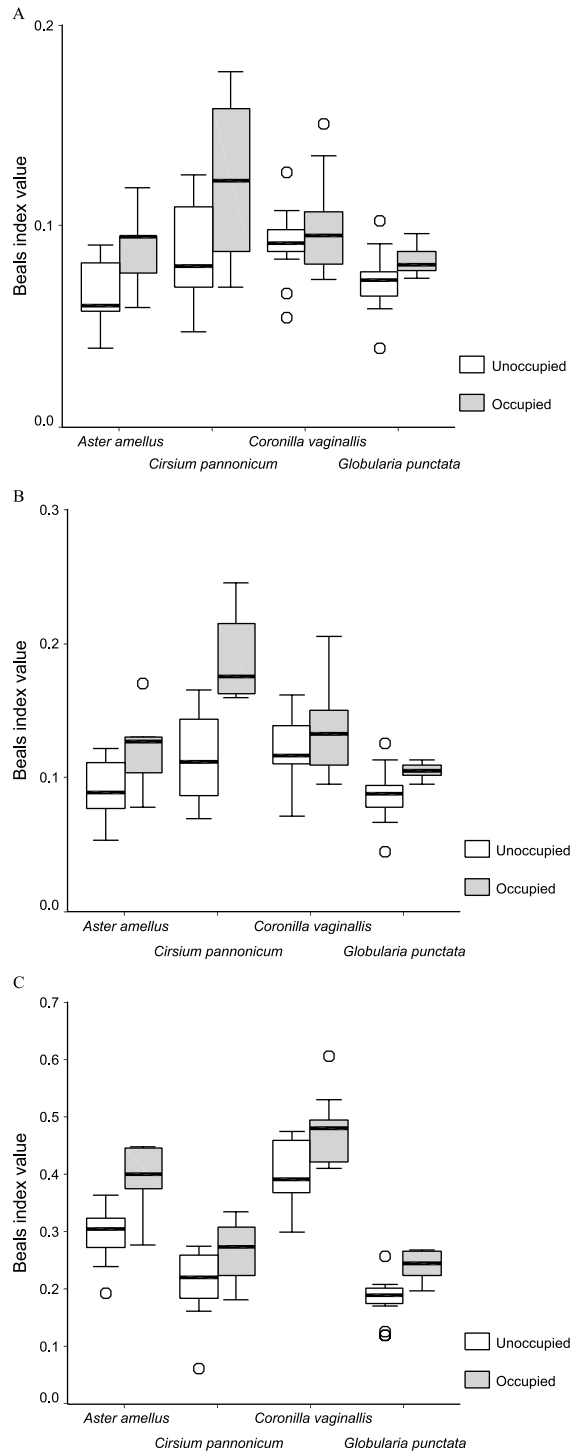


Fig. 2. Beals index values plotted separately for four species of dry grasslands. Values are separated for occupied and unoccupied habitats. The absolute values of Beals index depend on the frequency of the species in the database; therefore only comparisons within species make sense. Reference database: (A) all dry grasslands, (B) only *Bromion* dry grasslands, (C) the same database for which the prediction was done.

Table 1. Median Beals index values for actually occupied and unoccupied habitats calculated separately for each species. The values were calculated using either narrow or wide external reference database or without any external reference database (using data itself).

	Narrow		Wide		No	
	Unoccupied	Occupied	Unoccupied	Occupied	Unoccupied	Occupied
<i>Agrimonia eupatoria</i>	0.176	0.187	0.131	0.133	0.165	0.145
<i>Anemone sylvestris</i>	0.055	0.055	0.041	0.044	0.385	0.334
<i>Anthericum ramosum</i>	0.188	0.233	0.162	0.202	0.327	0.246
<i>Anthyllis vulneraria</i>	0.126	0.128	0.096	0.096	0.422	0.420
<i>Asperula cynanchica</i>	0.366	0.356	0.329	0.331	0.294	0.374
<i>Aster amellus</i>	0.089	0.127	0.060	0.095	0.304	0.401
<i>Astragalus onobrychis</i>	0.011	0.018	0.009	0.012	0.107	0.120
<i>Briza media</i>	0.243	0.270	0.200	0.233	0.117	0.081
<i>Bromus erectus</i>	0.208	0.254	0.157	0.182	0.106	0.049
<i>Carex humilis</i>	0.402	0.456	0.350	0.410	0.303	0.255
<i>Carex tomentosa</i>	0.034	0.036	0.029	0.033	0.348	0.288
<i>Carlina vulgaris</i>	0.257	0.262	0.184	0.189	0.593	0.579
<i>Centaurea scabiosa</i>	0.306	0.342	0.251	0.275	0.549	0.484
<i>Cirsium acaule</i>	0.370	0.404	0.256	0.294	0.759	0.761
<i>Cirsium eriophorum</i>	0.032	0.025	0.024	0.017	0.239	0.175
<i>Cirsium panonicum</i>	0.111	0.176	0.080	0.123	0.221	0.273
<i>Coronilla vaginialis</i>	0.116	0.132	0.092	0.095	0.391	0.481
<i>Eryngium campestre</i>	0.229	0.274	0.197	0.243	0.385	0.382
<i>Falcaria vulgaris</i>	0.077	0.087	0.067	0.080	0.207	0.198
<i>Galium verum</i>	0.291	0.376	0.250	0.313	0.228	0.240
<i>Geranium sanguineum</i>	0.100	0.133	0.099	0.146	0.105	0.070
<i>Globularia punctata</i>	0.088	0.105	0.073	0.081	0.190	0.245
<i>Helianthemum grandiflorum</i>	0.290	0.296	0.203	0.218	0.329	0.296
<i>Inula hirta</i>	0.073	0.088	0.050	0.058	0.207	0.174
<i>Knautia arvensis</i>	0.344	0.373	0.262	0.276	0.173	0.114
<i>Koeleria macrantha</i>	0.123	0.111	0.097	0.079	0.375	0.333
<i>Linum catharticum</i>	0.274	0.291	0.211	0.219	0.126	0.107
<i>Linum flavum</i>	0.063	0.083	0.043	0.057	0.239	0.258
<i>Linum tenuifolium</i>	0.111	0.129	0.083	0.097	0.359	0.302
<i>Melampyrum arvense</i>	0.031	0.030	0.034	0.034	0.169	0.090
<i>Ononis spinosa</i>	0.250	0.293	0.189	0.223	0.445	0.469
<i>Peucedanum cervaria</i>	0.097	0.131	0.076	0.100	0.272	0.209
<i>Plantago media</i>	0.407	0.422	0.304	0.320	0.337	0.317
<i>Primula veris</i>	0.097	0.115	0.082	0.106	0.269	0.269
<i>Prunella grandiflora</i>	0.180	0.225	0.137	0.172	0.596	0.575
<i>Salvia pratensis</i>	0.388	0.400	0.297	0.306	0.411	0.335
<i>Salvia verticillata</i>	0.122	0.108	0.097	0.087	0.366	0.289
<i>Scorzonera hispanica</i>	0.087	0.094	0.062	0.069	0.333	0.290
<i>Seseli hippomarathrum</i>	0.144	0.181	0.149	0.168	0.190	0.120
<i>Sesleria varia</i>	0.138	0.186	0.133	0.169	0.362	0.347
<i>Stachys recta</i>	0.141	0.164	0.161	0.187	0.363	0.371
<i>Tanacetum corymbosum</i>	0.109	0.150	0.090	0.119	0.406	0.368
<i>Tetragonolobus maritimus</i>	0.083	0.105	0.064	0.077	0.144	0.136
<i>Teucrium chamaedrys</i>	0.309	0.297	0.249	0.260	0.458	0.387
<i>Thesium linophylon</i>	0.040	0.056	0.038	0.058	0.196	0.166

reference database was used (Fig. 3). The relative width of the confidence intervals was larger for the percentiles of Beals index value than for the Beals index values themselves. This is due to two sources of stochasticity combined in the estimation of the percentiles – stochasticity of estimation of the Beals index values from occupied and unoccupied habitats.

Discussion

Patterns of Beals index values

There was a clear difference in the range of Beals index values estimated for occupied and unoccupied habitats.

Whereas most occupied habitats had a high Beals index value, unoccupied habitats had both high and low values for most of the species. This is an indication that the method used is able to differentiate between habitats and has therefore a predictive power to decide on habitat suitability. In case of several species there were a few occupied habitats having very low Beals index value. It can be speculated that these could be remnant populations, where the species survives just thanks to its longevity and is already not able to reproduce (Eriksson 1996). However other reasons, such as occurrence of the species on the margins of its ecological range or biased composition of the reference database, might also explain this pattern.

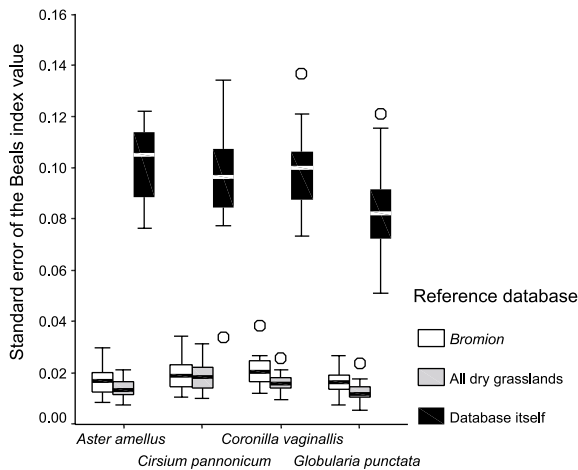


Fig. 3. Standard errors of Beals index values calculated by bootstrapping the database used to estimate species co-occurrence patterns. Two different external databases and the data itself were used for this purpose.

Predictions if the reference database is not available

Predictions without external reference database produced quite wide confidence intervals of the Beals index value. Still there was a clear difference between Beals index values estimated for unoccupied and occupied habitats and the prediction seems reasonable. The relationship between percentiles of Beals index value at unoccupied habitats estimated using external reference database and without it was however quite weak, even though there was quite good correspondence between the Beals index values. This shows that a rather small variation in Beals index values can cause quite a large variation in percentiles of Beals index value. This is due to a relatively low number of occupied habitats used to estimate the percentiles. It shows that even if external reference database is not available, one can gain quite reasonable estimates of the Beals index value. If the percentiles of Beals index values are to be calculated using this approach, one has to make sure that there are enough occupied habitats to construct a reliable cumulative frequency distribution. In our case they were too few.

Potentials and drawbacks of the method

The method presented here was originally developed for adding missing species to large databases. It worked with the assumption that species occupy all habitats they can and their absence in the database is just caused by not having been recorded. Their incidence in the database is therefore expected to be high. Metapopulation theory, however, predicts that not all possible species are present

at a habitat. The use of the method in the metapopulation context can therefore be limited depending on the fraction of habitats that are actually occupied; the lower the actual incidence of the species used for the prediction, the lower the possibility to make a good prediction. This brings a kind of contradiction into the issue, as one has to assume that the incidence of a species is reasonably high so that its occurrence patterns have predictive power, and still to expect a fraction of habitats to be unoccupied.

The width of the confidence interval of the Beals index values was quite wide even in case when the large external reference database was used. One could speculate that the width of the confidence interval will be a product of definition of external database (range of communities cover by the database) and range of species lists from the localities of interest. If both species lists are too wide, one may expect wide confidence intervals due to presence of a high number of species without any predictive power. On the other hand, if it is too narrow, the confidence intervals would be wide due to very few species used for the prediction. Our comparison of two different external databases however demonstrated that there is a very good correspondence between Beals index values estimated using these two and that also the confidence intervals of these two are comparable. This shows that selection of external database is not a major issue, and any reasonably large database can be used for this purpose. Detailed exploration of this issue is however beyond the scope of this paper.

This method is suggested as an easier to apply and possibly more informative counterpart to use of environmental variables. However given the longevity of many plant species, it may be argued that environmental factors such as soil conditions may be more informative when the environment had recently dramatically changed and the plants have not responded yet.

Conclusions

Our study provides a first insight into the possibilities to use species co-occurrence patterns to identify habitat suitability. Further exploration is needed to provide test of the method by comparing it with other approaches. Further work also needs to be done to evaluate the effect of various external database definitions and habitat definitions on the prediction.

Acknowledgements – We would like to thank Milan Chytrý for introducing to us the Beals index and for providing us with data from the phytosociological database and to Luboš Tichý for including Beals index calculation in the Juice program. We would also like to thank Jan van Groenendael and Johan Ehrlén for critical comments on the previous version of the manuscript and to Daniela Münzbergová for language revision. This study was supported by grants GACR no. 206/02/05, GAUK no. 156/

References

- Beals, E. W. 1984. Bray-Curtis Ordination: an effective strategy for analysis of multivariate ecological data. – *Adv. Ecol. Res.* 14: 1–55.
- Chytrý, M. and Rafajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. – *Preslia* 75: 1–15.
- Chytrý, M., Tichý, L., Holt, J. et al. 2002. Determination of diagnostic species with statistical fidelity measures. – *J. Veg. Sci.* 13: 79–90.
- Chytrý, M., Tichý, L. and Roleček, J. (in press). Local and regional patterns of species richness in Central European vegetation types along the pH/calcium gradient. – *Folia Geobot.* 38.
- Ehrlén, J. and Eriksson, O. 2000. Dispersal limitation and patch occupancy in forest herbs. – *Ecology* 81: 1667–1674.
- Eriksson, O. 1996. Regional dynamics of plants: a review of evidence for remnant, source-sink and metapopulations. – *Oikos* 77: 248–258.
- Eriksson, O. and Ehrlén, J. 1992. Seed and microsite limitation of recruitment in plant-populations. – *Oecologia* 91: 360–364.
- Ewald, J. 2002. A probabilistic approach to estimating species pools from large compositional matrices. – *J. Veg. Sci.* 13: 191–198.
- Freckleton, R. P. and Watkinson, A. R. 2002. Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. – *J. Ecol.* 90: 419–434.
- Gustafsson, C., Ehrlén, J. and Eriksson, O. 2002. Recruitment in *Dentaria bulbifera*; the roles of dispersal, habitat quality and mollusc herbivory. – *J. Veg. Sci.* 13: 719–724.
- Hanski, I. and Simberloff, D. 1997. The metapopulation approach, its history, conceptual domains, and application to conservation. – In: Hanski, I. and Gilpin, M. (eds), *Metapopulation biology*. Academic Press, pp. 5–26.
- Husband, B. C. and Barrett, S. H. 1996. A metapopulation perspective in plant population biology. – *J. Ecol.* 84: 461–469.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. – *Bull. Entomol. Soc. Am.* 15: 237–240.
- Losos, E. 1995. Habitat specificity of 2 palm species-experimental transplantation in Amazonian successional forests. – *Ecology* 76: 2595–2606.
- Ouborg, N. J. 1993. Isolation, population-size and extinction—the classical and metapopulation approaches applied to vascular plants along the Dutch Rhine-system. – *Oikos* 66: 298–308.
- Quintana-Ascencio, R. F. and Menges, E. S. 1996. Inferring metapopulation dynamics from patch-level incidence of Florida scrub plants. – *Conserv. Biol.* 10: 1210–1219.
- SPSS Inc. 2001. SPSS for Windows (Release 11.0.1), SPSS Inc.
- Turnbull, L. A., Crawley, M. J. and Rees, M. 2000. Are plant populations seed-limited? A review of seed sowing experiments. – *Oikos* 88: 225–238.
- Witte, J-P. M. 1998. National water management and the value of nature. – PhD Thesis, Agric. Univ., Wageningen.