# Environmental correlates of the patterns of plant distribution at the meso-scale: a case study from Northern Bohemia (Czech Republic)

**Faktory prostředí korelující s rozšířením rostlin na středním měřítku: případová studie ze severních Čech**

Petr Petřík & Jan Wild

*Institute of Botany, Academy of Sciences of the Czech Republic, CZ-252 43 Průhonice, Czech Republic; e-mail: petrik@ibot.cas.cz, wild@ibot.cas.cz*

Petřík P. & Wild J. (2006): Environmental correlates of the patterns of plant distribution at the meso-scale: a case study from Northern Bohemia (Czech Republic). – Preslia 78: 211–234.

The distribution of vascular plants in grid-cells and its relationship to the environmental correlates (driving factors) were studied using numerical methods (divisive classification and ordination). The first level of division in the classification distinguished forest and non-forest groups of grid-cells, and the second level four groups (containing predominantly species of base-rich forests at high altitudes, species of acidophilous mountain forests and small mountain grasslands, ruderal and meadows species at low altitudes, and species of thermophilous and basiphilous fringes and abandoned meadows). Within the study area, geographically consistent areas were delimited by correlating the groups, indicated by the divisive classification, with altitude and forest cover. Most differences in the Ellenberg indicator values for species in these groups for light, temperature, continentality, soil reaction and soil moisture were statistically significant. A number of variables were effective predictors (e.g. potential direct solar irradiation), physical geography (altitude, slope), land-cover (forest cover, area of urban zones) and geological bedrock were the key determinants of the species composition in the study area. However, even the most spatially correlated (according to Moran's I measure) were the naturally contiguous variables such as topographical features (altitude, slope and aspect). Generally, the grid-cells at low altitudes contained more species due to the co-occurrence of man-made habitats with fragments of semi-natural habitats. A relatively large percentage of the variation (15.8%) was accounted for by the spatial structure of the data, the environmental factors explained 18.9%, but 65.3% of the floristic variance remained unexplained. The most spatially autocorrelated variables were also the most correlated with regard to species composition. However, the relatively high autocorrelation in the species data and their derivates had comparable or lower effect on species composition than the most autocorrelated environmental factors. The results were compared with those of other European studies, and possible bias due to the different ways of collecting and analysing data, and effect of different scales discussed.

Keywords: biogeography, Ellenberg indicator values, flora, grid mapping, spatial autocorrelation, variance partitioning

## Introduction

Traditionally, phytogeographical studies (particularly those based on distribution atlases) deal with the delimitation of biogeographical zones and definition of the chorological elements and species area ranges (Hultén & Fries 1986). A numerical approach to phytogeography has developed in recent years as a consequence of the huge amount of data in distribution atlases that has been digitized (see e.g. Schölzel et al. 2002). For identifying a repeatable pattern researchers use either a multivariate analysis (Lepš & Šmilauer 2003) or classification methods (Hill 1979). Many phytogeographers have attempted to predict plant occurrence and explain patterns in species distribution in terms of environmental factors, using a geographical information system (GIS), which links together ecological and statistical models of plant distribution (Collingham et al. 2000, Guisan & Zimmermann 2000, Vaughan & Ormerod 2003).

The number of fine-resolution studies that cover large geographical areas is still inadequate (but see Duda et al. 1990) because of the time it takes to collect the data. To circumvent the need for a large sampling effort, Gaston (1996) suggested an alternative method for assessing species richness based on indicator groups (surrogacy method). The question is how to define such groups. One approach is to use Ellenberg's indicator values for species ("Zeigerwerte" by Ellenberg et al. 2001), which are used in the present study. The indicator species (a special case of functional response groups, see Lavorel & Garnier 2002) are successfully used at various spatial scales assuming that their features correspond with certain environmental factors (Orschied 1994, von Numers & van der Maarel 1998, Chytrý et al. 1999, Korsch 1999).

The biogeographical data depend on many environmental variables, which are naturally spatially structured (Legendre & Legendre 1998, Legendre et al. 2002). Particularly, the arranging of samples in a grid template leads to an undesirably higher similarity of neighbouring grid-cells. The strong spatial structure of the data can influence the interpretation of the species–environment relationships. This topic is discussed in the huge literature on methods of spatial statistics (using e.g. local quadrate variance, autologistic models, block quadrate variance, correlograms, variograms, angular correlations, directional variograms, wavelets, nearest neighbour methods and various landscape ecology metrics, see Perry et al. 2002 or Dale & Fortin 2002 for reviews). For the detection of large spatial trends in species distribution, variance partitioning (partial CCA) has recently become widely used, due to the availability of ordination methods in CANOCO software (Borcard et al. 1992, Lepš & Šmilauer 2003). The aim of this procedure is to split the total floristic variance into spatially and non-spatially structured parts by including the spatial position of the grid-cells in the analysis. An analysis of spatial autocorrelation is possible by examining the correlation of variables within the data set as a function of distance in correlograms using Moran's I measure (Moran 1950, Upton & Fingleton 1988). In this study both variance partitioning and correlograms were used to show the spatial dependence of the data.

The aims of the present study are: (1) to determine whether species composition shows consistent associations at the meso-scale; (2) to assess the relationships between environmental factors (gradients) and species distribution; (3) to use the indicator values for the indirect assessment of the influence of the given factors on species composition; (4) to assess the component of the floristic variance due to spatial effects. Meso-scale data were recorded using a grid of cells of 0.52 $km^2$ in an area of nearly 100 $km^2$, in the predominantly forested Ještědský hřbet mountain range (Czech Republic).

## Materials and methods

### *Study area*

The field survey was carried out in the Ještědský hřbet mountain range (N Bohemia) and neighbouring area (Fig. 1). The range extends in a NW–SE direction (between 50°40'–50°49' N and 14°52'–15°03' E) along a geological fault of sediments poor in minerals in the west (sandstone) and granite with Quaternary sediments in the east, with a geo-

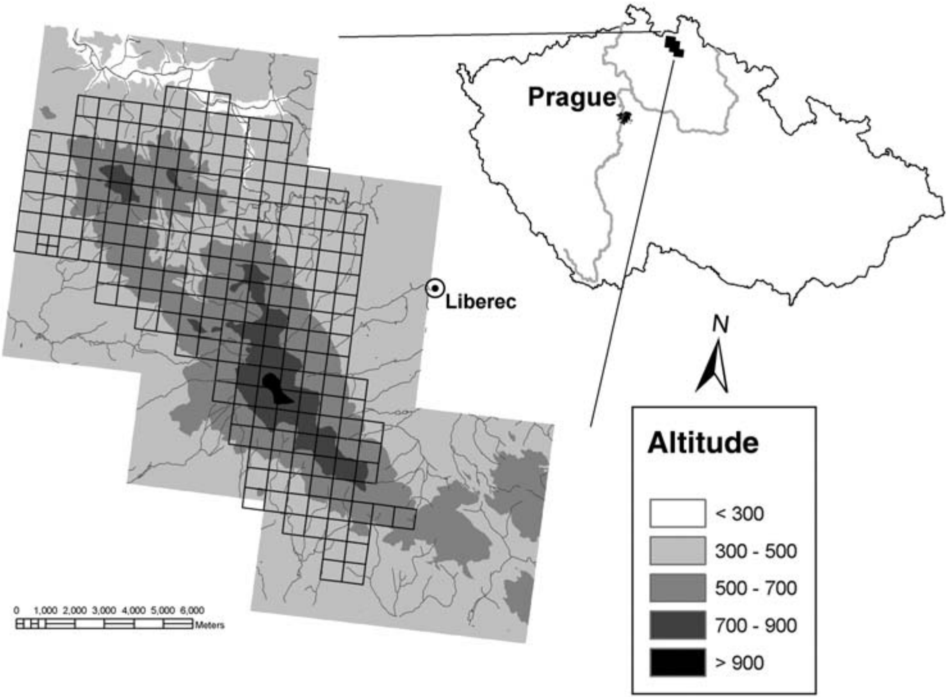

Fig. 1. – Study area superimposed on a grid of 192 cells (ca 0.52 km$^2$) of 5/8 × 3/8 degree minutes. Altitude is in metres.

logically complicated part in the centre of the mountain range consisting of different types of metamorphic bedrock, plutonic rock and sediments.

Micro- and meso-climatic conditions may vary substantially due to the broad range of altitudes (270–1012 m a.s.l.) and the heterogeneity of the terrain, with a mean annual temperature of 4 (near the summits) to 7 °C. The annual precipitation is between 800–1000 mm (Hostýnek 1993), and humid air blows down from the top of Ještěd. Hence, the most humid conditions prevail at altitudes of about 800 m a.s.l. The Lužická Nisa river (and its tributary Rokytka) are the only large rivers that flow through the NE margin of the area.

The northern and central parts of the study area are colder and drier, and the flora characteristically consists of forest species with a suboceanic distribution, while thermophilous continental elements are typical of the warmer SE part. According to the phytogeographical national division (Skalický 1988), the study area is a part of the mesophilous floristic district Mesophyticum, which is typical of mid altitudes, and has a lower floristic diversity. The actual vegetation cover and land use consist of a mosaic of managed meadows, *Picea abies* plantations and the remnants of beech forest.

### *Floristic survey*

The presence of native plant species and naturalized and invasive aliens (see Pyšek et al. 2004 for definitions) was recorded systematically within the defined grid-cells from 1998 to 2004. Casual aliens, agricultural crops and planted ornamentals were not recorded. For mapping, the Central European Basic Area (CEBA) system was used so that the results could be compared with those of previous studies from Central Europe. The system is based on grid template divided into squares of 10 × 6 arc minutes (in Germany it is the "Messtischblatt" – MTB) unlike the Universe Transverse Mercator (UTM) system, which is based on a kilometre grid and is used mainly in W and N Europe. Each grid-cell (5/8 × 3/8 arc minutes of the CEBA, i.e. ca 740 × 700 m = 0.52 km$^2$) was visited at least twice, once in early spring and once in summer. The time spent on a cell depended on the environmental heterogeneity of that cell and was between one and two days. Some cells, mainly marginal ones, which were sampled with less intensity, were excluded from the analysis. In addition, 47 very rare species were excluded because they might have been planted (e.g. *Salix alba*). To account for uneven field sampling and minimize identification bias, 23 species were merged as aggregates (e.g. *Agrostis stolonifera* agg.). As a consequence, the data set used for numerical analysis contained 843 species and 192 grids with 828 species containing information on environmental variables.

The data were stored in a TURBOVEG for Windows 2.0 database (Hennekens & Schaminée 2001). The nomenclature of species follows Kubát et al. (2002).

### *Data sets with environmental factors*

The digital topographic model (DTM) was derived from a map with 10-m line contours (scale 1:10,000) provided by the Czech Office for Surveying, Mapping and Cadastre. Variables derived from the DTM were intersected by the 192 grid-cells using GIS software ArcView 3.2 (ESRI 1999) with the Spatial Analyst extension. Minimum, mean, and maximum values were calculated for: altitude, aspect, slope and Potential Direct Solar Irradiation (PDSI hereafter). The PDSI describes how much incident radiation a grid-cell receives during a certain period and is calculated from slope and aspect of the terrain, taking into account shading of the grid-cell by the horizon (Conrad 2002). The PDSI values were estimated for every hour of every seventh day of the vegetation period (from March to October) and summed for each grid-cell using the programme DiGeM (Conrad 2002), which was applied to the DTM.

The land-cover information came from a map of scale 1:25,000 provided by the Military Geographical and Hydrometeorological Institute and the following digitized schematic layers were used for a land-cover assessment within each grid-cell: the total surface of urban areas, the total length of roads and railways, the total area of forests, clearings, and water bodies, the total length of shrub, river (river-basin > 5 m) and stream (basin < 5 m) lines. To avoid the effect of the close correlation between several selected environmental variables, similar variables were combined. This was the case for the network of linear shrubby vegetation structures such as avenues, windbreaks, and hedgerows, which are often accompanied by natural shrubs (combined as shrub line). The 86 geological bedrock types present in the study area were reduced to eight main units based on their chemical and physical properties: granite and acid volcanic rock; metaconglomerate; quartzite and silicate; metamorphosed pelite (i.e. phyllite and mica schist); base-rich volcanic rock and metamorphic rock (i.e. green schist and amphibolite); Quaternary sediments (i.e. alluvium and delluvium); chalk and marl; loam and loess. The bedrock map (1:50,000) is that of the Czech Geological Survey.

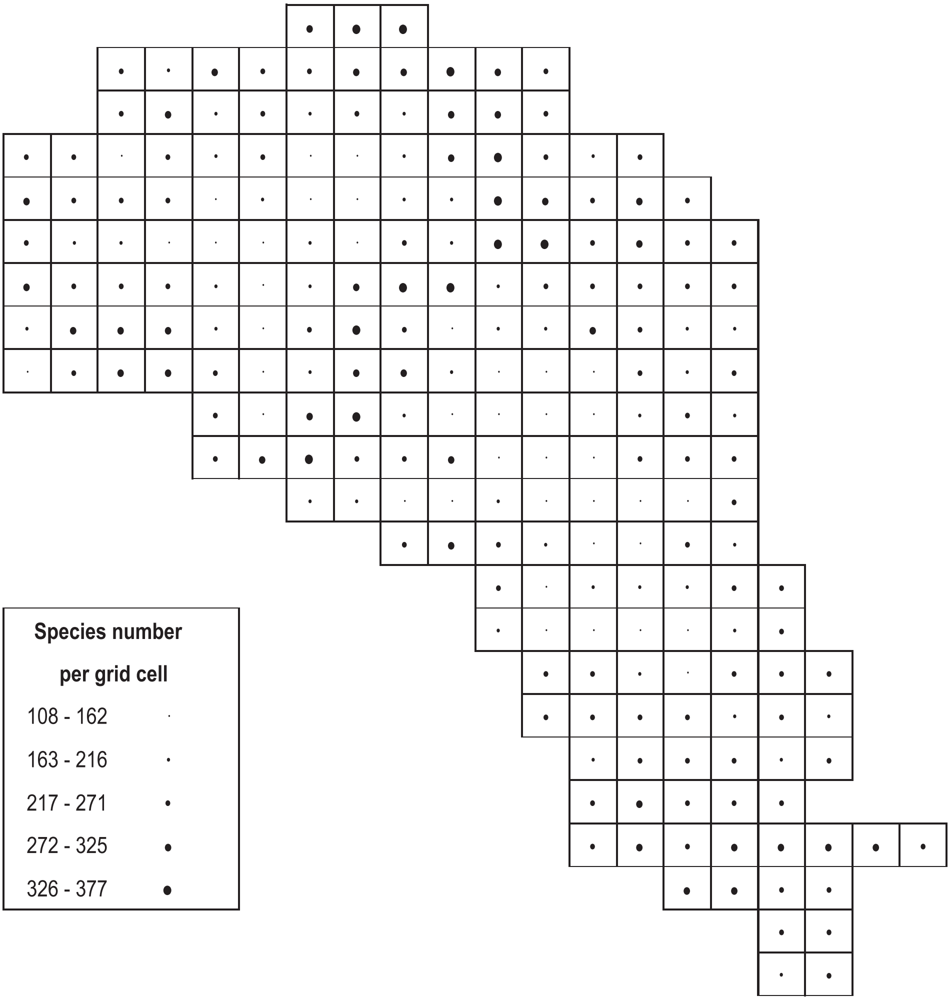

Fig. 2. – Map of species richness (α diversity) in the study area. The circle sizes correspond to the ranges in species richness.

Additional information was extracted from the species data set and used as supplementary variables in constrained ordinations: species number, Ellenberg indicator values (Ellenberg et al. 2001) for light (L), soil moisture (F), temperature (T), soil nutrients (N; as a measure of productivity), soil reaction (R; high total calcium content; Schaffers & Sýkora 2000) and continentality (K). In our data set, 8.4% of species lacked an indicator value because of their wide ecological amplitude. Values were also unavailable for neophytes, cultivated species and species with a natural distribution outside the range covered by Ellenberg's study (e.g. *Petasites kablikianus*). This was the case for values for temperature (missing for 32% species); 21% of species lacked values for continentality, soil moisture, soil reaction and soil nutrients, respectively, and 10% for light. The descriptive statistics of all the 26 explanatory and seven supplementary variables are given in Appendix 1.

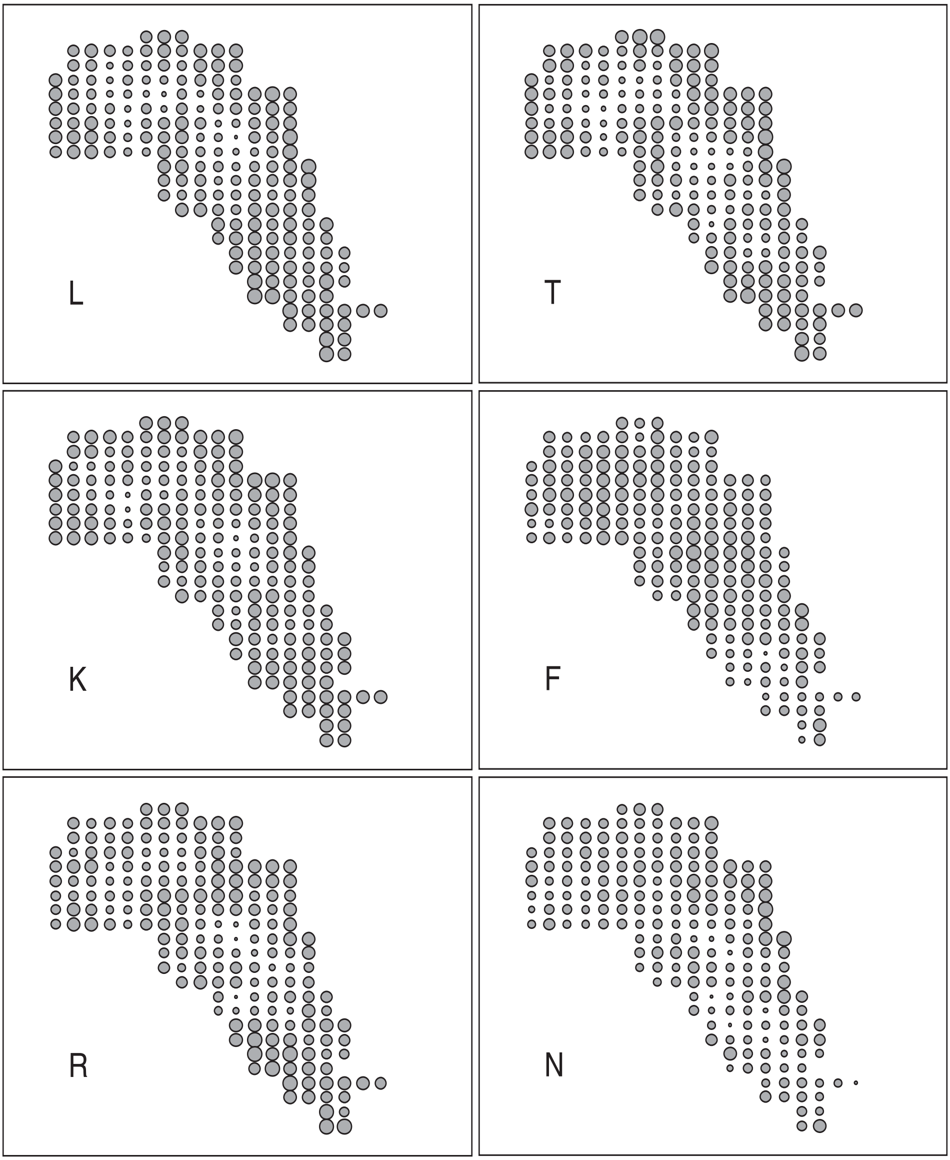

Fig. 3. – Geographical pattern of average Ellenberg indicator values in the 192 grid-cells. L – light; T – temperature; K – continentality; F – soil moisture; R – soil reaction (total calcium); N – soil nitrogen (productivity). The circle sizes accord with the range of values.

The variables present in Appendix 1 are those used after merging the "redundant" ones. Since correlated variables interfere with each other and may lead to problems in model selection, six variables that were highly intercorrelated (correlation coefficient ≥ 0.65) were excluded from the analysis.

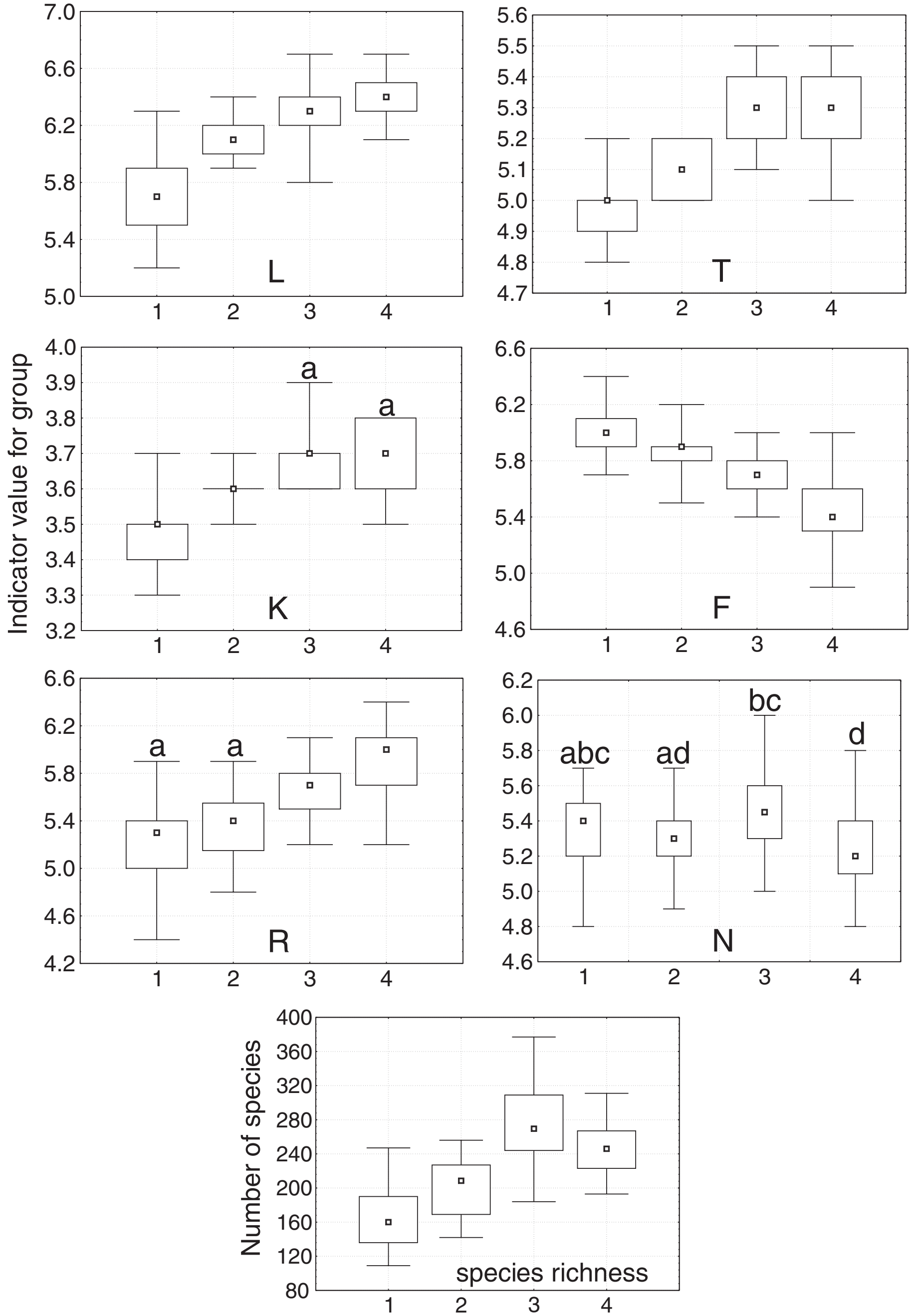

Fig. 4. – Box-plot diagrams for the TWINSPAN groups 1–4 for indicator values and species richness. L – light; T – temperature; K – continentality; F – soil moisture; R – soil reaction (total calcium); N – soil nitrogen (productivity). Open small squares indicate medians, boxes interquartile ranges, whiskers non-outlier ranges. Boxes with the same letter were not statistically different at $P < 0.05$. Post hoc comparison by Tukey honest significant difference test for unequal size (Spjotvoll/Stoline test) was used for multiple comparisons.

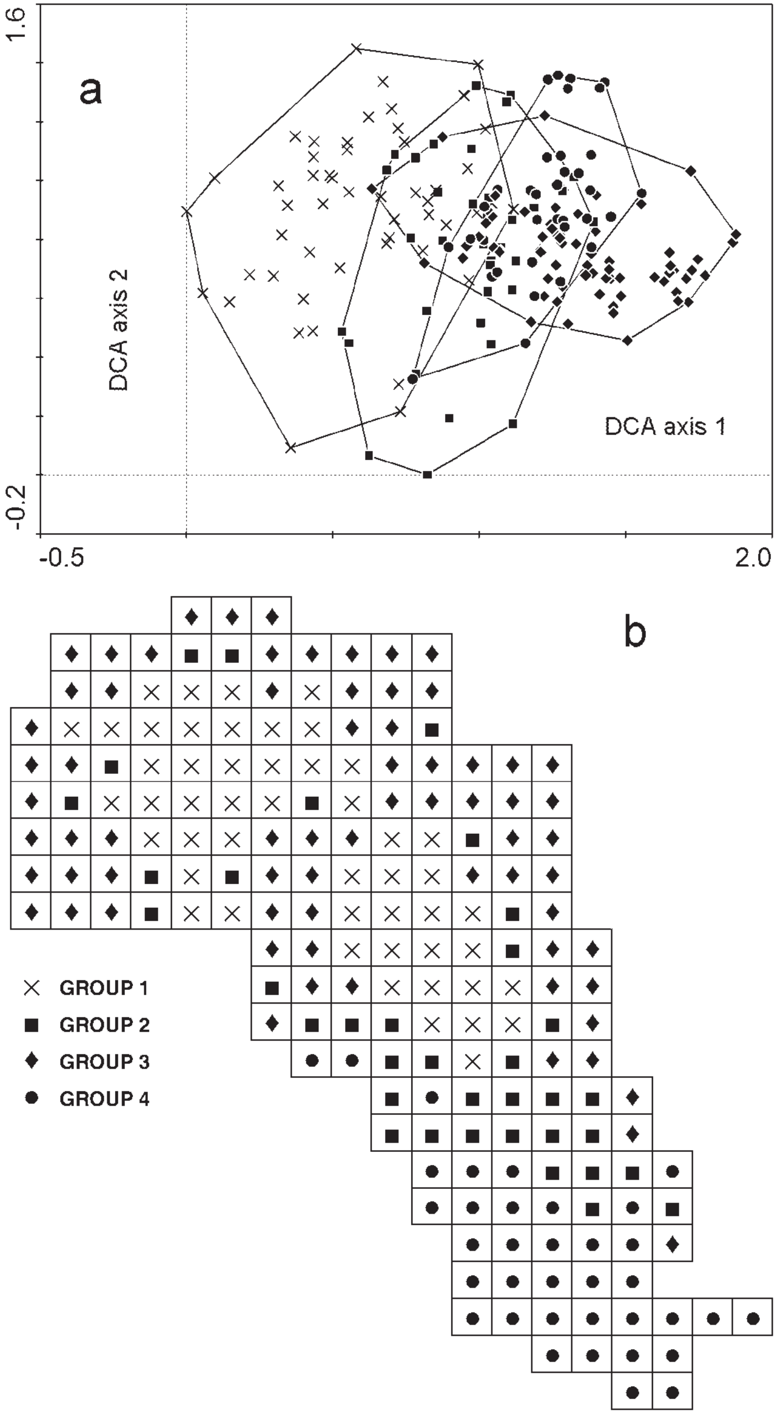

Fig. 5. – TWINSPAN classification results of the distribution data shown as (a) DCA-diagram and (b) geographical pattern of the groups.

### *Classification of species composition*

A hierarchical classification, two-way indicator species analysis TWINSPAN (Hill 1979) in programme JUICE 6.3 (Tichý 2002) was used to find floristically similar grid-cells (with a maximum of six levels of division and a minimum group size of five). To analyse the floristic composition of the resulting classification units, a synoptic table was constructed. In this table, the diagnostic species of the individual groups were determined using the phi coefficient (Sokal & Rohlf 1995, Chytrý et al. 2002). The phi coefficient was adjusted for equal size of clusters (Tichý & Chytrý in prep.). The size of all groups was standardized to a size equal to 25% of the total data set using a Fisher's exact test and a significance level $P < 0.05$. Ellenberg indicator values were calculated for each classification group. For statistical comparisons of the differences between the groups, the software package STATISTICA 5.1 (StatSoft Inc. 1998) was used.

### *Ordination*

Unimodal-based techniques in CANOCO for Windows 4.0 (ter Braak & Šmilauer 2002) were used to discern trends in the composition of species. For the unconstrained ordination, Detrended Correspondence Analysis (DCA) with geographical coordinates as covariates (to filter out spatial autocorrelation) was chosen. The default options according to Lepš & Šmilauer (2003) (i.e. detrending by segments, no data transformation, downweighting rare species) were followed.

For the assessment of species composition–environment correlations (i.e. the data set with environmental variables and that containing species occurrences), a Canonical Correspondence Analysis (CCA) with geographical coordinates as covariates was performed. In the CCA, inter-species distances and other steps identical to those in the DCA were followed. To find the minimum number of statistically significant variables, a manual forward selection procedure with 999 Monte Carlo significance permutation tests was used.

The eigenvalues and percentages of floristic variance explained by the first four DCA and CCA axes were calculated. The pattern obtained from the classification was transferred onto a graph with sample groups marked in the DCA. The geographical pattern of scores of the first two canonical axes is shown for the CCA.

### *Spatial pattern analyses*

To reveal and remove spatial trends in the distribution data the geographical coordinates of grid-cells and their transformation reflecting different trend surfaces were used. First, all grid-cells were coded in the geographical coordinate system x and y. Second, seven additional geographical variables were derived from these coordinates by adding their 2nd and 3rd polynomial terms and their order interactions and together they were used as explanatory variables in polynomial regression (Borcard et al. 1992). In this way not only the linear but also the quadratic and cubic distributions of the data were described. All nine polynomial terms were statistically significant (permutation tests carried out in the CCA using forward selection). The data sets and geographical variables were subjected to variance partitioning with four partially constrained ordinations (Borcard et al. 1992). The total floristic variance was thus reduced to: (a) purely spatial variance; (b) spatially structured environmental variation; (c) non-spatial environmental variance; and (d) unexplained variance. Accordingly, four partial CCAs were performed. These analyses were constrained by: (1) significant environmental factors only; (2) all nine forms of coordinates (tested by permutation tests); (3) environmental variables with coordinates as covariates; and (4) coordinates with environmental variables as covariates.

To explore the intensity of the spatial pattern in both environmental variables (only average values) and species composition (DCA scores), spatial autocorrelations expressed as correlograms of Moran's I, was used. Rook's adjacency of grid-cells (four cardinal neighbours) was defined in the regular lattice data providing coefficients (I) for ten lags based on mutually exclusive distance classes in correlograms. Test of significance of lags was tested based on 999 permutations (programme produced by J. W.).

## Results

During the mapping survey, a total of 1082 taxa (including hybrids) were recorded. Of these, 153 are on the Red List of the Czech Republic (Holub & Procházka 2000) and 42 are protected by Czech law. Forty-nine previously reported threatened taxa were not re-found during the mapping (probably due to successional changes in their habitats and inappropriate landscape management, see also Petřík & Višňák 2006). The number of species per grid-cell varies between 108 and 377 (Fig. 2). The richest grid-cells are those located at low altitudes and the poorest located at high altitudes and in forested areas.

The geographical pattern in Ellenberg indicator values seems to be associated with altitude and forest cover (land use) but not for soil moisture and nutrients (Fig. 3). The values for soil nitrogen decrease from the eastern urbanized to the western forested and rural landscape and those for moisture show a S–N trend. The characteristics of classification groups based on indicator values is given in Fig. 4.

### *Classification of species composition at the meso-scale*

Within the first division of the classification the forest and non-forest groups were distinguished (Table 1). Further classification of these groups was ecologically best interpreted at the second level of division, which resulted in four main species groups. The species in the first group are herbs of species-rich ravines and beech forests at mid altitudes or forest springs. The second group is less clearly defined (due to low fidelity values) and consists of more acidophilous and montane species, although species with similar habitat requirements appear also in the first group (e.g. *Lastrea limbosperma*, see also the Discussion). In the third and best-defined group (with the highest phi values), there are species of arable land (e.g. *Chenopodium polyspermum*) and mesophilous meadows at low altitudes (e.g. *Sanguisorba officinalis*, not shown in Table 1). The thermophilous and basiphilous species are concentrated in the fourth group.

The differences between the classification groups are significant (except the nutrient values) when indicator values are used (for reaction values particularly). From the first to the fourth species group, which demand more light and a continental climate, indicator values for temperature, continentality and soil reaction increase (Fig. 4). There is a similar trend in species richness, with the exception of a high value in the third group. On the other hand, the soil moisture decreases.

The groups at the second level of classification are only partially separable along to the first two axes of the DCA ordination (Fig. 5a). The first four DCA axes explain 17.6% of

Table 1. – Species groups derived from the TWINSPAN classification of distribution data. Only two levels of divisions are highlighted with only 10 characteristic species for each group (with the number of grid-cells and ranges of fidelity values).

| No. of group | Group 1 | Group 2 |
|---|---|---|
| No. of grid-cells | 90 | 123 |
| Fidelity range | 0.56–0.42 | 0.83–0.71 |
| First division | *Phegopteris connectilis* | *Fallopia convolvulus* |
| | *Cardamine flexuosa* | *Viola arvensis* |
| | *Polygonatum verticillatum* | *Chenopodium album* agg. |
| | *Lysimachia nemorum* | *Festuca pratensis* |
| | *Calamagrostis villosa* | *Veronica persica* |
| | *Carex remota* | *Arabidopsis thaliana* |
| | *Chrysosplenium oppositifolium* | *Holcus lanatus* |
| | *Digitalis purpurea* | *Phleum pratense* |
| | *Prenanthes purpurea* | *Lolium perenne* |
| | *Circaea ×intermedia* | *Veronica arvensis* |

| No. of group | Group 1 | Group 2 | Group 3 | Group 4 |
|---|---|---|---|---|
| No. of grid-cells | 45 | 45 | 84 | 39 |
| Fidelity range | 0.51–0.33 | 0.38–0.22 | 0.63–0.47 | 0.75–0.61 |
| Second division | *Melica uniflora* | *Petasites albus* | *Hieracium sabaudum* | *Cornus sanguinea* |
| | *Gymnocarpium dryopteris* | *Calamagrostis canescens* | *Oxalis fontana* | *Cruciata laevipes* |
| | *Hordelymus europaeus* | *Trientalis europaea* | *Calystegia sepium* | *Euphorbia cyparissias* |
| | *Milium effusum* | *Juncus squarrosus* | *Chenopodium polyspermum* | *Origanum vulgare* |
| | *Festuca altissima* | *Juncus bulbosus* | *Poa palustris* | *Centaurea scabiosa* |
| | *Circaea ×intermedia* | *Euphrasia stricta* agg. | *Arctium lappa* | *Fragaria moschata* |
| | *Arum maculatum* | *Salix aurita* | *Symphoricarpos albus* | *Ranunculus bulbosus* |
| | *Prenanthes purpurea* | *Juncus filiformis* | *Verbascum nigrum* | *Carlina acaulis* |
| | *Lastrea limbosperma* | *Spergularia rubra* | *Senecio vulgaris* | *Convolvulus arvensis* |
| | *Dentaria enneaphyllos* | *Ranunculus flammula* | *Sonchus oleraceus* | *Briza media* |

Table 2. – Interset correlations of the first two canonical axes in CCA and results of the forward selection (linear, quadratic, and triple combinations of coordinates were used as covariates). Only the first unit of bedrock (chalk and marl) was tested for significance due to inter-correlation with other units, which were then automatically included into the model. Numbers in shaded fields > |0.25|. F – F statistic values, significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

| Variable | CCA 1 | CCA 2 | F |
|---|---|---|---|
| Forest cover (area of forest) | 0.713 | 0.272 | 9.67*** |
| Average slope | 0.245 | 0.679 | 5.01*** |
| Average altitude | 0.721 | 0.013 | 4.74*** |
| Length of river | –0.141 | 0.230 | 2.00*** |
| Area of urban zones | –0.478 | –0.086 | 1.47*** |
| Average PDSI | –0.250 | –0.085 | 1.45** |
| Area of clearings | 0.396 | 0.028 | 1.37** |
| Length of streams | –0.404 | 0.030 | 1.35** |
| Chalk and marl | –0.083 | 0.488 | 2.35*** |
| Granite | 0.052 | –0.010 | – |
| Metaconglomerate | 0.279 | 0.056 | – |
| Silicite | 0.139 | –0.113 | – |
| Metamorphosed pelite | 0.249 | 0.173 | – |
| Base-rich volcanic rocks | 0.027 | 0.192 | – |
| Quaternary sediments | –0.412 | –0.272 | – |
| Loam and loess | –0.090 | –0.254 | – |

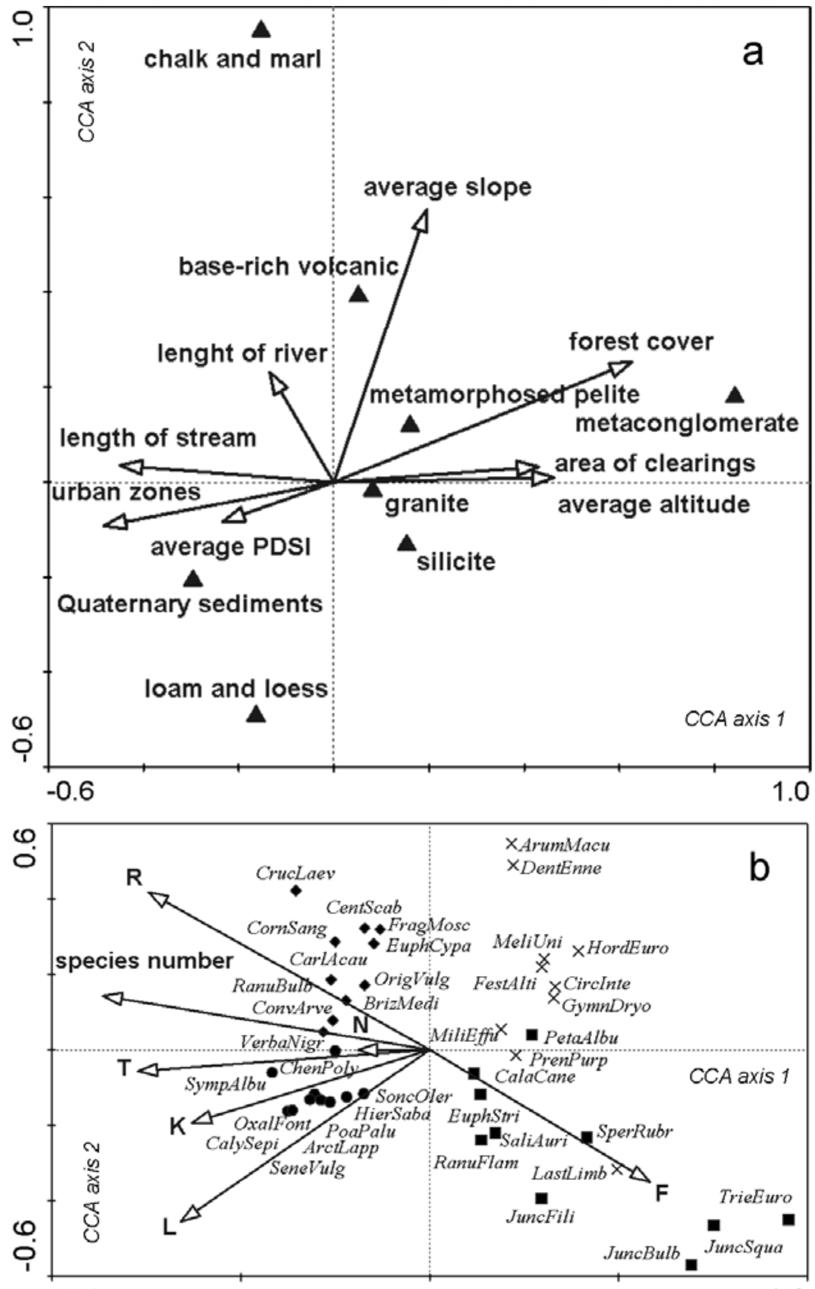

Fig. 6. – Ordination diagrams depicting the first two CCA axes. (a) Projection of significant environmental variables chosen by forward selection (the categorial variables are indicated by triangles rather than arrows). (b) Projection of supplementary variables with species of TWINSPAN groups. Names of taxa (four first characters of generic and specific names) abbreviated according to Table 1. For abbreviations of the variables, see Appendix 1 and for symbols of the species groups, see Fig. 5.

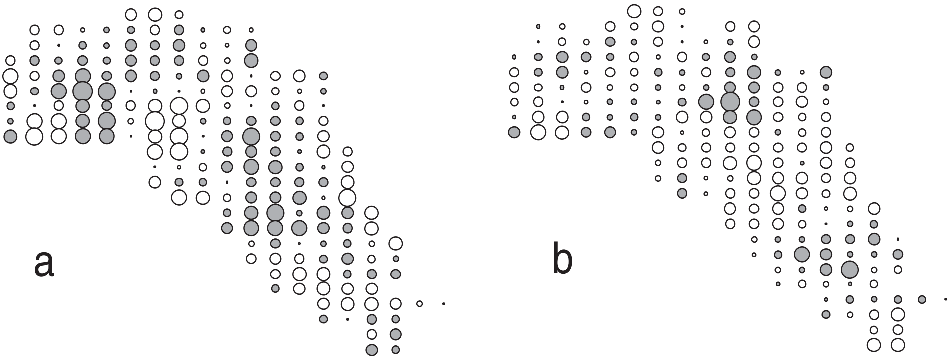

Fig. 7. – Geographical pattern of CCA scores from the first (a) and second (b) canonical axis. Sample scores are based on linear combinations of environmental variables in CANOCO. Circle sizes correspond to the ordination scores (positive scores filled circles, negative open).

the variance due to the non-random spatial distribution of classification groups. The percentages of variances explained by the first, second, third and fourth axes are: 9.1%, 4.4%, 2.2%, and 1.9%. For all the classification groups an obvious geographical pattern exists (Fig. 5b). While the first two groups are concentrated at forested areas of middle and high altitudes, the third group is at the margins of the entire area; this group reaches the central northern part of the study area (Rokytka stream valley). The fourth group is limited to the southern part of the study area and has a distribution pattern similar to that for moisture.

*Species composition–environment relationship*

The first four CCA ordination axes explain 12.5% of the variability in the distribution of species. The first two axes capture nearly three quarters (71%) of the variability of the first two axes of the unconstrained DCA ordination. The first CCA axis explains 6.7% of the variance and the second 3.2%. The other two axes are difficult to interpret as they account for very little of the floristic variance (1.4% and 1.2%, respectively). Of the 20 explanatory variables used, 16 passed successfully through the forward selection procedure (Table 2). Average aspect, the occurrences of shrubs, ponds, and the length of roads and railways were not chosen by the forward selection. The first canonical axis (with the highest eigenvalue) is positively correlated with the area of forests and clearings, average altitude and the occurrence of metaconglomerates (see also Fig. 6a). The urban zones, Quaternary sediments, length of streams and average PDSI are negatively correlated with the first axis. The second axis' strongest positive correlation is with the average slope, chalk and marl, length of river and forest cover. The second axis is most negatively correlated with Quaternary sediments and loam and loess. The distributions of the diagnostic species revealed by TWINSPAN classification show clear correlations with the variables (Fig. 6b). These results correspond well with the geographical pattern of sample scores, based on linear combinations of environmental variables in CANOCO (Lepš & Šmilauer 2003), of the first two constrained canonical axes (Fig. 7).

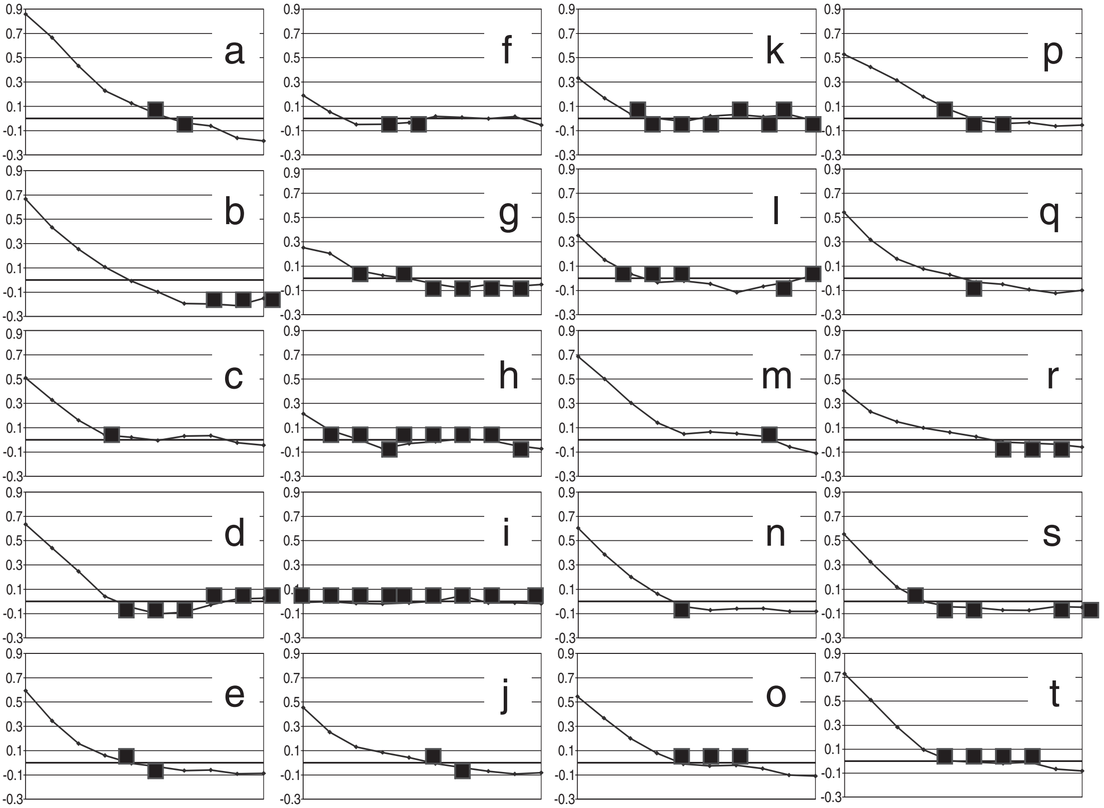

Fig. 8. – Moran's correlograms of the 12 variables used in the multivariate analysis. The values of Moran's I (y-axes) ranging from –0.3 to 0.9 were calculated for 10 lags (x-axes). The geology of the bedrock was not included as it is a categorial variable. (a) average altitude; (b) average aspect; (c) average slope; (d) average PDSI; (e) area of forests; (f) length of shrub lines; (g) area of clearings; (h) length of streams; (i) area of ponds; (j) length of river; (k) extent of urban areas; (l) length of roads and railways; indicator values for (m) light, (n) temperature, (o) continentality, (p) soil moisture, (q) soil reaction, and (r) soil nutrients; (s) average species number; (t) scores of the first DCA axis. The black boxes indicate insignificant values at P = 0.05 (999 permutation tests).

### *Spatial pattern analyses*

The total variance of species assemblages can be partitioned as follows: (a) spatial variance not shared by environmental variables (8.0%); (b) spatially structured environmental variance (7.8%); (c) pure environmental variance (18.9%); (d) unexplained variance (65.3%). Thus, space accounts for about 45% of the variability that is due exclusively to the environment.

The strong spatial autocorrelation of environmental variables is revealed by Moran's I correlograms (Fig. 8). It usually decreases from the first lag (i.e. grid-cell size = 0.7km) with increasing distance between lags. The most autocorrelated (with Moran's I on the first lag between 0.65 and 0.85) and significant up to the fourth to seventh lag are topographical variables (altitude, PDSI and aspect). Similarly, slope, area of forests and presence of river (first lag: I = 0.45–0.60, positively autocorrelated up to fourth lag) have relatively high values and are particularly dependent on the physical geography. Quite unimportant is the spatial autocorrelation between urban areas and roads (first lag about: I = 0.35, significant up to second lag). Almost without a spatially defined pattern (first lag: I ≤ 0.20, significant up first to third lag) are ponds, clearings, minor streams and shrub lines. However, up to the fourth lag (ca 3 km), the autocorrelation rapidly decreases and then only fluctuates along the null values. The comparison of various variables derived from the species data set (i.e. species richness and Ellenberg values) shows relatively high spatial autocorrelations at the first lag and, generally, the shape of the correlogram curves is comparable with that of the most autocorrelated environmental factors (Fig. 8m–r). Similarly, the Moran's I correlogram of DCA scores (Fig. 8t) indicates that vegetation patch size is comparable to or lower than that of the most autocorrelated environmental factors.

## Discussion

### *Consideration on species and species indicators*

The classification of species using either a divisive technique or an ordination method is satisfactory as both approaches gave comparable results. The geographical distribution of samples of the classified groups (in fact, their species composition) is interpretable in terms of the drivers used in constrained ordination. The diagnostic species of the classification groups resulted in coherent groups in ordination space. There are only two exceptions in the first group: *Lastrea limbosperma* and *Petasites albus*. This may be because the occurrence of these species is associated with forest paths; thus, they may transgress into various environments along viatic corridors.

Although the composition of species associations (species groups) can be different at different spatial scales, some species groups were spatially stable in the study area (e.g. the groups of herb-rich and ravine forests; Petřík & Bruelheide 2006). The species of such spatially stable groups are included in the first classification group, however, this group is not very consistent in ordination space. The reason may be the various methods used (Cocktail vs. ordination). Besides the method used, the stability of a species group's composition depends on gradients in a study area (Kuželová & Chytrý 2004). For, example, in contrast to that obtained in the very heterogeneous landscape of the Bavarian Alps (Schmidtlein & Ewald 2003), there was no positive correlation between Ellenberg indicator value for soil reaction (correlating with the occurrence of chalk and marl) and altitude in this study. This is probably due to the scattered occurrence of calcium-rich substrates in the Czech study area.

The classification of species groups (grid-cells) used in the present study is only valid for the local scale of the study and it can be assumed that it would change with increasing size of the study area. This local classification is determined ecologically while the regional (e.g. nationwide) one geographically (Knollová & Chytrý 2004). However, the geographical approach is not useful for the local scale (see Knollová & Chytrý 2004). In addition, the aim of the classification in this study was not to establish new phytosociological units but describe within-regional floristic variability.

As indicator values show optimal conditions rather than tolerance, Schaffers & Sýkora (2000) suggest different data adjustment for indicator values. On the other hand, Käfer & Witte (2004) do not recommend any averaging. Furthermore, the unevenness in indicator values (see Methods) could influence the results only in a species-poor data set (Chytrý et al. 1999) and, in addition, the Ellenberg indicator values are not very sensitive to the completeness of data (Ewald 2003).

### *Consideration of selected predictors*

At scales similar to the present study, floristic variance is relatively well reflected by primary environmental factors (i.e. parameters related to resource availability; Heikkinen & Birks 1996, Pausas & Austin 2001). These scales favour geological, topographical and biotic factors at the expense of climatic factors. Many biogeographical studies highlight the fact that climatic factors seem to be the most important variables operating at large scales (Haeupler 1974, Birks 1976, Preston & Hill 1997, Dahl 1998, Storch et al. 2003, Thuiller et al. 2004). However, such fine scale data are generally not available and therefore authors use various surrogates. In the present study, potential direct solar irradiation was the only broad climate-related variable used.

Considering the constraints mentioned below, in most European studies, variables connected with altitude, slope, bedrock, or land-cover (Appendix 2) are found to be the best predictors of species distribution. There are many other studies that partially confirm our results. For instance, for the correlation of species richness with native and alien plant distribution, Kühn et al. (2003) found geological diversity to be the only correlate taking a spatial autocorrelation into account. Similarly, Deutschewitz et al. (2003) attributed most of the variability in the distribution of both natives and plant aliens to riverine and urban ecosystems. Fédoroff et al. (2005) concluded that species diversity is dependent on landscape heterogeneity and disturbance, which accords with our results (i.e. a low species richness in homogeneous forest and undisturbed sites). Both at low and high altitudes there are species-poor habitats such as arable fields and spruce plantations, respectively. Arable fields are highly disturbed compared to undisturbed forest plantations. At low altitudes, however, a combination of fine-scale landscape structures and intermediate disturbances in grasslands can result in a high species diversity. The role of the grasslands, located mainly at low altitudes, supports the intermediate disturbance hypothesis (see Rejmánek et al. 2004).

Important driving factors were demonstrated to vary across scales (Pearson & Dawson 2003); these factors should be studied at various scales (Levin 1992) or, at least, at the scale appropriate to the process observed (Austin 2002). In the study area, however, all the environmental variables considered affected the species composition in the same way, regardless of the grain of study (Petřík & Bruelheide 2006). Therefore, the results of the species composition–environment correlation may be applied at various scales.

Environmental variables can be classified as indirect, direct or resource gradients (Austin 2002). The indirect gradients represent, in our case, topographical features (e.g. altitude), and their correlation with plant distribution due to their high spatial-dependence and connection to the temperature (i.e. direct gradient). Resource variables (e.g. nutrient supply) in the present study were replaced by indicator values. These surrogates are “the collective properties” of Austin (2002), who further distinguished between proximal and distal environmental predictors (gradients) based on the position of the predictor in the chain of impact of processes between plant and predictor. Robust distribution models should be based on proximal (i.e. causal variable determining the plant growth) and direct variables

(Austin 2002), however, it is unlikely a digital map for proximal variables will be constructed (e.g. a map of soil nutrient solutions).

Fortunately, resource and direct variables can be replaced by indirect variables, such as altitude, which is a highly complex variable related to energy parameters including changes in temperature, rainfall, wind, etc. However, this inter-correlation makes tests of hypotheses associated with the altitudinal gradient difficult (Rahbek 2005). For instance, the general decrease of species richness with altitude could be accounted for in terms of energy supply (Wright et al. 1993) and at a narrow range of altitudes the species richness–altitude relationship tends to be linear instead of unimodal (Rahbek 2005). The same is true in the present study, however, one must bear in mind that the acidic bedrock and large *Picea abies* plantations depauperized the flora at high altitudes in the study area.

The generally significant relationship between the occurrence of road networks and species distribution (Gelbard & Belnap 2003, Pauchard & Alaback 2004, Hansen & Clevenger 2005) was absent from our analysis. This might be due to the temporal, geometric, or thematic resolution of the layers used in this analysis (e.g. digitizing the road network from existing topographic maps did not distinguish between road use as was the case in Parendes & Jones 2000). The cited papers highlight the importance of altitude and land-use (i.e. forested areas serve as barriers against invasion by alien species). It seems that the same variables reduce the spread of aliens in the study area, however, the important role of forest clearings (as a place for potential invasion of forests, see e.g. Prach et al. 1995) may enhance the invasion by aliens; clearings are relatively rich in alien species (Chytrý et al. 2005). Again, low thematic resolution might be responsible for the insignificance of the occurrences of shrubs and ponds in the analysis.

### *Spatial dependence of data*

The most spatially correlated are variables that are naturally contiguous and dependent on the terrain such as altitude or PDSI, respectively. On the other hand, the variables with scattered, linear, or of rare occurrence are not spatially structured. The grain mosaic in the species data is comparable to that of the variables with the highest spatial dependence. Thus, patches of species are not formed by factors other than those used in analysis. This indicates that the grain of the present study was appropriate for studying environmental factors that operate at the meso-scale.

The most spatially autocorrelated variables are those best correlated with species composition. These strong spatial autocorrelations in our data set could result in inflated type I error rate of the test on the species composition–environment relationship (Legendre et al. 2002). However, some phytogeographical studies ignore the spatial dependence of grid data, which might be unimportant (Heikkinen & Birks 1996, Heikkinen et al. 1998, Korvenpää et al. 2003) or the main explanatory factor of species variance in the data set (Storch et al. 2003).

### *Constraints of comparison with similar studies*

The comparison of driving factors obtained from grid-mapping is difficult for two reasons: species' patterns and attributes of scale are scale-dependent (Kunin 1998, Scheiner et al. 2000) and the explanatory variables used at a fine scale are not applicable at a broad scale and vice versa because of the lack of data appropriate for a given scale (Wiens 1989, Levin

1992, Rahbek 2005). In addition, local factors operating at fine spatial and temporal scales cannot reveal the species–environment relation if the resolution is too broad (Wiens 1989).

Recording effort and data processing influence any interpretation. In many phytogeographical studies either common or rare species are excluded from the analysis (usually all species with > 95% and < 5% occurrence; Pedersen 1990, Myklestad 1993) and the use of an unrestricted data set is rare (but see Chytrý et al. 1999). The geographical shape of a sampled area influences the structure of the data set (especially of islands; von Numers & van der Maarel 1998). The distribution (in terms of species density) in grid-cells may not represent the real occurrence (ecological density), as areas with no occurrences are usually also included (Gaston 1996). This problem results from uneven recording effort, which biases species data sets (Rich & Woodruff 1992, Rich & Smith 1996, Petřík & Boublík 2003). Such bias is of relatively little importance in this study as only one person mapped the study area and there was adequate seasonal coverage. On the other hand, a systematic sampling design makes the grid data suitable for statistical analyses, compared to subjectively located relevés. Thus, a stratified gradsect sampling design, used in recent vegetation studies (e.g. Lososová 2004), is not needed in analyses of grid-mapping data.

In similar studies, it is not unusual for more than a half of the floristic variance to remain unexplained (Økland 1999). This could be due to polynomial distortion of the ordination method (Økland 1999) or permanent shifts in populations etc., but it is impossible to assess the real weight of the individual factors. The problem of seeking for causality in environmental gradients and of circularity associated with the use of indicator species was discussed e.g. by Wilson et al. (2004). In the present study, these problems in gradient analyses was resolved by using as many species as possible, broadly defined indicator values, and independent robust statistical methods.

## Acknowledgements

We thank Jan Tonika for his assistance with the geological map. We are grateful to Phil Lambdon, Zuzana Münzbergová, Petr Pyšek, Lubomír Tichý, and two anonymous referees for their valuable comments on the manuscript, to Jakub Lepš for improving the English of an earlier version of the paper, and to Tony Dixon for editing the final version. This research was partly supported by Biodiversity Research Centre (grant no. LC 06073 funded by Ministry of Education, Youth and Sports of the CR), grant 206/05/0020 (Czech Science Foundation) and no. AV0Z60050516 (Grant Agency, Academy of Sciences of the Czech Republic), and by BioHab (http://www.biohab.alterra.nl).

## Souhrn

Prostorové uspořádání dat ze síťového mapování cévnatých rostlin a jejich korelace s faktory prostředí jsme zkoumali pomocí numerických metod (divizivní klasifikace v programu TWINSPAN a ordinace v programu CANOCO). Klasifikace rozdělila mapovací pole nejdříve do dvou skupin podle zastoupení lesních a nelesních druhů. Druhé dělení proběhlo do čtyř skupin (s převažujícími druhy na živiny náročných listnatých lesů středních poloh, acidofilních horských lesů a maloplošných travních společenstev, ruderálních a lučních porostů nižších poloh a konečně termofilních a bazifilních lemů a lad). Jasně se ukázaly geograficky vymezené oblasti založené na korelaci skupin z divizivní klasifikace s gradientem nadmořské výšky a lesnatosti. Indikační hodnoty pro druhy z jednotlivých skupin potvrdily statisticky rozdílné nároky druhů na světlo, teplotu, kontinentalitu, půdní reakci, a půdní vlhkost; nikoliv však na živiny v půdě. Variabilita druhového složení mapovacích polí byla v přímé gradientové analýze (CCA s postupným členěním variability) nejlépe vysvětlena pomocí topografie (nadmořská výška, sklonitost), land-cover (lesnatosti a plochy zastavěného území), geologického podkladu a přítomnosti větší

řeky (Lužické Nisy). Obecně platilo, že druhově nejbohatší byla území nižších poloh díky kombinaci zastoupení lidmi ovlivněných biotopů se zbytky polopřirozené vegetace. Relativně velkou část variability (15,8 %) vysvětlila prostorová struktura v datech, zatímco 18,9 % vysvětlily použité proměnné prostředí, ale 65,3 % variability zůstala nevysvětlena. Nejvíce prostorově korelovány byly právě proměnné s kontinuálním rozložením jako jsou nadmořská výška, lesnatost, sklonitost a orientace. Nicméně velká prostorová závislost se projevila také v druhových datech a v údajích z nich odvozených (skóry DCA, druhová bohatost, indikační hodnoty) ve srovnatelné nebo v menší míře, než tomu bylo u nejvíce autokorelovaných proměnných prostředí. Tato silná autokorelace v použitých databázích mohla vést ke snížené výpovědní hodnotě testů ve vztahu druhové složení–faktory prostředí. Výsledky jsme srovnali s ostatními evropskými studiemi, avšak bez snahy o větší generalizaci kvůli problémům spojeným s rozdílným sběrem, analýzou dat a vzhledem k závislosti následné interpretace na použitém měřítku.

## References

Austin M. P. (2002): Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. – Ecol. Model. 157: 101–118.

Birks H. J. B. (1976): The distribution of European pteridophytes: a numerical analysis. – New Phytol. 77: 257–287.

Borcard D., Legendre P. & Drapeau P. (1992): Partialling out the spatial component of ecological variation. – Ecology 73: 1045–1055.

Chytrý M., Grulich V., Tichý L. & Kouřil M. (1999): Phytogeographical boundary between the Pannonicum and Hercynicum: a multivariate analysis of landscape in the Podyjí/Thayatal National Park, Czech Republic/Austria. – Preslia 71: 23–41.

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., Tichý L., Holt J. & Botta-Dukát Z. (2002): Determination of diagnostic species with statistical fidelity measures. – J. Veg. Sci. 13: 79–90.

Collingham Y. C., Wadsworth R. A., Huntley B. & Hulme P. E. (2000): Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. – J. Appl. Ecol. 37 (Suppl. 1): 13–27.

Conrad O. (2002): DiGem 2.0. – URL [http://www.geogr.uni-goettingen.de/pg/saga/digem/index.html]

Dahl E. (1998): The phytogeography of northern Europe (British Isles, Fennoscandinavia and adjacent areas. – Cambridge University Press, Cambridge.

Dale M. R. T. & M.-J. Fortin (2002): Spatial autocorrelation and statistical tests in ecology. – Écoscience 9: 162–167.

Deutschewitz K., Lausch A., Kühn I. & Klotz S. (2003): Native and alien plant species richness in relation to spatial heterogeneity on a regional scale in Germany. – Glob. Ecol. Biogeogr. 12: 299–311.

Duda J., Herben T. & Novotný J. (1990): Correlations of liverwort species on an intermediate landscape scale. – J. Veg. Sci. 1: 623–628.

Ellenberg H., Weber H. E., Düll R., Wirth V. & Werner W. (2001): Zeigerwerte von Pflanzen in Mitteleuropa. Ed. 3. – Scr. Geobot. 18: 1–258.

ESRI (1999): ArcView GIS. – Environmental Systems Research Institute Inc. Redlands.

Ewald J. (2003): The sensitivity of Ellenberg indicator values to the completeness of vegetation relevés. – Basic Appl. Ecol. 4: 507–513.

Fédoroff É., Ponge J.-F., Dubs F., Fernández-González F. & Lavelle P. (2005): Small-scale response of plant species to land-use intensification. – Agr. Ecosyst. Environ. 105: 283–290.

Gaston K. J. (ed.) (1996): Biodiversity. A biology of numbers and difference. – Blackwell Science, London.

Gelbard J. L. & Belnap J. (2003): Roads as conduits for exotic plant invasions in a semiarid landscape. – Conserv. Biol. 17: 420–432.

Guisan A. & Zimmermann N. E. (2000): Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147-186.

Hansen M. J. & Clevenger A. P. (2005): The influence of disturbance and habitat on the presence of non-native plant species along transport corridors. – Biol. Conserv. 125: 249–259.

Haeupler H. (1974): Statistische Auswertungen von Punktrasterkarten der Gefäßpflanzenflora Süd-Niedersachsens. – Scr. Geobot. 8: 1–141.

Heikkinen R. K. (1996): Predicting patterns of vascular plant species richness with composite variables: a meso-scale study in Finnish Lapland. – Vegetatio 126: 151–165.

Heikkinen R. K. (1998): Can richness patterns of rarities be predicted from mesoscale atlas data? A case study of vascular plants in the Kevo Reserve. – Biol. Conserv. 83: 133–143.

Heikkinen R. K. & Birks H. J. B. (1996): Spatial and environmental components of variation in the distribution patterns of subarctic plant species at Kevo, N Finland – a case study at the meso-scale level. – Ecography 19: 341-351.

Heikkinen R. K., Birks H. J. B. & Kalliola R. J. (1998): A numerical analysis of the mesoscale distribution patterns of vascular plants in the subarctic Kevo Nature Reserve, northern Finland. – J. Biogeogr. 25: 123–146.

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

Hill M. O. (1979): TWINSPAN. A Fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. – Cornell University, Ithaca.

Holub J. & Procházka F. (2000): Red list of vascular plants of the Czech Republic – 2000. – Preslia 72: 187-230.

Hostynek J. (1993): Meteorologie, klimatologie a hydrologie. Úvodní klimatická studie – Ještědský hřeben. – In: Herejk J. (ed.), Ještědský hřeben. Vstupní inventarizační studie, p. 40–67, Liberec.

Hultén E. & Fries M. (1986): Atlas of North European vascular plants north of the Tropic of Cancer, vols. I–III. – Koeltz Scientific Books, Königstein.

Käfer J. & Witte J.-P. M. (2004): Cover-weighted averaging of indicator values in vegetation analyses. – J. Veg. Sci. 15: 647–652.

Knollová I. & Chytrý M. (2004): Oak-hornbeam forests of the Czech Republic: geographical and ecological approaches to vegetation classification. – Preslia 76: 291–311.

Korvenpää T., von Numers M. & Hinneri S. (2003): A mesoscale analysis of floristic patterns in the south-west Finnish Archipelago. – J. Biogeogr. 30: 1019–1031.

Korsch H. (1999): Chorologisch-ökologische Auswertung der Daten der Floristischen Kartierung Deutschlands. – Schriftenr. Vegetationskd. 30: 1–200.

Kubát K., Hrouda L., Chrtek J. jun., Kaplan Z., Kirschner J., Štěpánek J. & Zázvorka J. (2002): Klíč ke květeně České republiky. – Academia, Praha.

Kühn I., Brandl R., May R. & Klotz S. (2003): Plant distribution patterns in Germany – Will aliens match natives? – Feddes Repert. 114: 559–573.

Kunin W. (1998): Extrapolating species abundance across spatial scales. – Science 281: 1513–1515.

Kuželová I. & Chytrý M. (2004): Interspecific associations in phytosociological data sets: how do they change between local and regional scale? – Plant Ecol. 173: 247–257.

Lavorel S. & Garnier E. (2002): Predicting changes in community composition and ecosystem functioning from plant traits: revising the Holy Grail. – Funct. Ecol. 16: 545–556.

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.

Lepš J. & Šmilauer P. (2003): Multivariate analysis of ecological data using Canoco. – Cambridge University Press, Cambridge.

Levin S. (1992): The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.

Lososová Z. (2004): Weed vegetation in southern Moravia (Czech Republic): a formalized phytosociological classification. – Preslia 76: 65–85.

Moran P. A. P. (1950): Notes on continuous stochastic phenomena. – Biometrika 37: 17–23.

Moser D., Dullinger S., Englisch T., Niklfeld H., Plutzar C., Sauberer N., Zechmeister H. G. & Grabherr G. (2005): Environmental determinants of vascular plant species richness in the Austrian Alps. – J. Biogeogr. 32: 1117–1127.

Myklestad A. (1993): The distribution of *Salix* species in Fennoscandia – a numerical analysis. – Ecography 16: 329–344.

Orschied O. (1994): Auswertung der Rasterkartierung pfälzischer Gefäßpflanzen in Bezug auf Zeigerwerte. – Mitt. Pollichia 81: 313–327.

Økland R. H. (1999): On the variation explained by ordination and constrained ordination axes. – J. Veg. Sci. 10: 131–136.

Parendes L. A. & Jones J. A. (2000): Role of light availability and dispersal in exotic plant invasions along roads and streams in the H. J. Andrews experimental forest, Oregon. – Conserv. Biol. 14: 64–75.

Pauchard A. & Alaback P. B. (2004): Influence of elevation, land use, and landscape context on patterns of alien plant invasions along roadsides in protected areas of South-Central Chile. – Conserv. Biol. 18: 238–248.

Pausas J. G. & Austin M. P. (2001): Patterns of plant species richness: an appraisal. – J. Veg. Sci. 12: 153–166.

Pausas J. G., Carreras J., Ferré A. & Font X. (2003): Coarse-scale plant species richness in relation to environmental heterogeneity. – J. Veg. Sci. 14: 661-668.

Pearson R. G. & Dawson T. P. (2003): Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? – Glob. Ecol. Biogeogr. 12: 361–371.

Pedersen B. (1990): Distributional patterns of vascular plants in Fennoscandia: a numerical approach. – Nord. J. Bot. 10: 163–189.

Perry J. N., Liebhold A. M., Rosenberg M. S., Dungan J., Miriti M., Jakomulska A. & Citron-Pousty S. (2002): Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. – Ecography 25: 578–600.

Petřík P. & Boublík K. (2003): Sources of variation in botanical grid mapping. – Novit. Bot. Univ. Carol. 17: 17–23.

Petřík P. & Bruelheide H. (2006): Species groups can be transferred across different scales. – J. Biogeogr. 33 (doi: 10.1111/j. 1365–2699.2006.01514.x)

Petřík P. & Višňák R. (2006): Zur Flora und Vegetation des Jeschkenkammes. – Ber. Naturforsch. Ges. Oberlausitz 14 (in press)

Prach K., Hadinec J., Michálek J. & Pyšek P. (1995): Forest planting as a way of species dispersal. – Forest Ecol. Manag. 76: 191–195.

Preston C. D. & Hill M. O. (1997): The geographical relationships of British and Irish vascular plants. – Bot. J. Linn. Soc. 124: 1–120.

Pyšek P., Richardson D. M., Rejmánek M., Webster G., Williamson M. & Kirschner J. (2004): Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. – Taxon 53: 131–143.

Rahbek C. (2005): The role of spatial scale and the perception of large-scale species-richness patterns. – Ecol. Lett. 8: 224–239.

Rejmánek M., Rejmánková E. & Holzner W. (2004): Species diversity of plant communities on calcareous screes: the role of intermediate disturbance. – Preslia 76: 207–222.

Rich T. C. G. & Smith P. A. (1996): Botanical recording, distribution maps and species frequency. – Watsonia 21: 161–173.

Rich T. C. G. & Woodruff E. R. (1992): Recording bias in botanical surveys. – Watsonia 19: 73–95.

Schaffers A. P. & Sýkora K. V. (2000): Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. – J. Veg. Sci. 11: 225–244.

Scheiner S. M., Cox S. B., Willig M., Mittelbach G. G., Osenberg C. & Kaspari M. (2000): Species richness, species–area curves and Simpson's paradox. – Evol. Ecol. Res. 2: 791–802.

Schmidtlein S. & Ewald J. (2003): Landscape patterns of indicator plants for soil acidity in the Bavarian Alps. – J. Biogeogr. 30: 1493–1503.

Schölzel C. A., Hense A., Hübl P., Kühl N. & Litt T. (2002): Digitization and geo-referencing of botanical distribution maps. – J. Biogeogr. 29: 851–856.

Skalický V. (1988): Regionálně fytogeografické členění. – In: Hejný S. & Slavík B. (eds.), Květena ČSR 1, p. 103–121, Academia, Praha.

Sokal R. R. & Rohlf F. J. (1995): Biometry. Ed. 3. – Freeman, New York.

StatSoft Inc. (1998): STATISTICA (data analysis software system), version 5.1. – URL [http://www.statsoft.cpm]

Storch D., Konvička M., Beneš J., Martínková J. & Gaston K. J. (2003): Distribution patterns in butterflies and birds of the Czech Republic: separating effects of habitat and geographical position. – J. Biogeogr. 30: 1195–1205.

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. – Microcomputer Power, Ithaca.

Thuiller W., Araujó M. B. & Lavorel S. (2004): Do we need land-cover data to model species distribution in Europe? – J. Biogeogr. 31: 353–361.

Tichý L. (2002): JUICE, software for vegetation classification. – J. Veg. Sci. 13: 451–453.

Upton G. J. G. & Fingleton B. (1988): Spatial data analysis by example. Vol. I – Point pattern and quantitative data. – John Wiley & Sons, Chichester.

Vaughan I. P. & Ormerod S. J. (2003): Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. – Conserv. Biol. 17: 1601–1611.

von Numers M. & van der Maarel E. (1998): Plant distribution patterns and ecological gradients in the Southwest Finnish archipelago. – Global Ecol. Biogeogr. 7: 421–440.

Wiens J. (1989): Spatial scaling in ecology. – Funct. Ecol. 3: 385–397.

Wilson J. B., Agnew A. D. Q. & Sykes M. T. (2004): Ecology or mythology? Are Whittaker's "gradient analysis" curves reliable evidence of continuity in vegetation? – Preslia 76: 245–253.

Wright D. H., Currie D. J. & Maurer B. A. (1993): Energy supply and patterns of species richness on local and regional scales. – In: Ricklefs R. E. & Schluter D. (eds.), Species diversity in ecological communities: historical and ecological perspectives, p. 66–73, The University of Chicago Press, Chicago.

Received 18 October 2005
Revision received 5 April 2006
Accepted 16 April 2006

Appendix 1. – Variables used in multivariate analysis. * – variable excluded as the correlation coefficient with other variables was significantly higher than 0.65 at P = 0.05. PDSI = potential direct solar irradiation.

| Description | Units | Min. | Average | Max. | Std. dev. |
|---|---|---|---|---|---|
| Topography and geography | | | | | |
| Minimum altitude* | m a.s.l. | 279.3 | 455.4 | 833.9 | 109.9 |
| Average altitude | m a.s.l. | 294.7 | 529.6 | 908.9 | 128.6 |
| Maximum altitude* | m a.s.l. | 352.0 | 611.5 | 1,010.0 | 141.0 |
| Average aspect | degrees | 36.4 | 160.8 | 281.7 | 62.2 |
| Minimum slope* | degrees | 0.0 | 0.8 | 6.8 | 1.2 |
| Average slope | degrees | 2.9 | 12.9 | 23.1 | 4.7 |
| Maximum slope* | degrees | 6.1 | 36.6 | 58.3 | 9.8 |
| Bedrock geology | | | | | |
| Area of granite | $m^2$ | 0.0 | 39,267.0 | 482,675.4 | 99,242.4 |
| Area of metaconglomerate | $m^2$ | 0.0 | 9,912.3 | 203,390.3 | 35,200.1 |
| Area of silicite | $m^2$ | 0.0 | 75,459.0 | 503,083.0 | 109,207.6 |
| Area of metamorphosed pelite | $m^2$ | 0.0 | 178,563.8 | 511,036.3 | 160,078.4 |
| Area of base-rich volcanic rock | $m^2$ | 0.0 | 29,166.4 | 357,756.7 | 72,408.4 |
| Area of Quaternary sediments | $m^2$ | 0.0 | 135,410.1 | 511,347.4 | 144,380.0 |
| Area of chalk and marl | $m^2$ | 0.0 | 20,029.1 | 291,807.0 | 49,106.8 |
| Area of loam and loess | $m^2$ | 0.0 | 24,042.1 | 426,915.3 | 72,348.9 |
| Climate-related variable | | | | | |
| Minimum PDSI* | $kW \cdot m^{-2}$ | 170.9 | 717.6 | 1,314.6 | 251.1 |
| Average PDSI | $kW \cdot m^{-2}$ | 1,066.1 | 1,320.5 | 1,539.2 | 95.8 |
| Maximum PDSI* | $kW \cdot m^{-2}$ | 1,379.9 | 1,608.3 | 1,659.1 | 64.2 |
| Land-cover variables | | | | | |
| Area of forests (forest cover) | $m^2$ | 0.0 | 355,728.5 | 511,548.7 | 172,610.7 |
| Length of shrub lines | m | 0.0 | 185.6 | 3,245.6 | 443.7 |
| Area of clearings | $m^2$ | 0.0 | 179.7 | 1,831.0 | 326.3 |
| Length of streams | m | 0.0 | 496.6 | 1,753.3 | 448.2 |
| Area of ponds | $m^2$ | 0.0 | 36.5 | 2,438.2 | 252.4 |
| Length of river | m | 0.0 | 601.8 | 18,008.2 | 2,320.1 |
| Surface of urban areas | $m^2$ | 0.0 | 2,295.8 | 30,324.0 | 4,310.9 |
| Length of roads, railways | m | 555.6 | 3,091.3 | 6,451.2 | 1,207.2 |
| Collective properties | | | | | |
| Number of species | | 108 | 227.4 | 377 | 59.1 |
| Indicator value for light (L) | | 5.2 | 6.1 | 6.7 | 0.3 |
| Indicator value for temperature (T) | | 4.8 | 5.2 | 5.5 | 0.2 |
| Indicator value for continentality (K) | | 3.3 | 3.6 | 3.9 | 0.1 |
| Indicator value for soil moisture (F) | | 4.9 | 5.8 | 6.4 | 0.3 |
| Indicator value for soil reaction (R) | | 4.4 | 5.5 | 6.4 | 0.4 |
| Indicator value for soil nutrients (N) | | 4.8 | 5.4 | 6.0 | 0.2 |

Appendix 2. – Overview of the variables used to study plant species–environment correlations in European studies based on grid mapping. Variables which were insignificant or used only once were excluded. References: 1 – Myklestad (1993); 2 – Heikkinen (1996); 3 – Heikkinen & Birks (1996); 4 – Heikkinen (1998); 5 – Heikkinen et al. (1998); 6 – von Numers & van der Maarel (1998); 7 – Chytrý et al. (1999); 8 – Luoto (2000); 9 – Korvenpää et al. (2003); 10 – Pausas et al. (2003); 11 – Moser et al. (2005); 12 – the present study. Significance levels: * P < 0.05, ** P < 0.01, *** P < 0.001; Tested but non-significant factor: –.

| Used variable in the study | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Topography and geography | | | | | | | | | | | | |
| Minimum altitude (lowest point) | ** | ** | ** | *** | | ** | | | | – | | |
| Average altitude | | * | – | *** | | ** | | * | | – | | *** |
| Maximum altitude (highest point) | ** | – | ** | *** | | ** | *** | | | – | | |
| Altitudinal range | ** | | | *** | | | *** | | | | – | |
| Longitude | ** | ** | ** | | | | | | ** | | | *** |
| Latitude | ** | ** | ** | | | | | | ** | | | *** |
| Main aspect of slopes (average aspect) | | * | ** | | | – | | *** | | | | *** |
| Horizontal length of the high cliffs | | | | *** | | ** | | | | | | |
| Area of lakes (water bodies) | | – | ** | ** | | ** | | | | | | – |
| Length of rivers | | ** | ** | *** | | ** | | | | | | *** |
| Length of rivers and brooks (streams) | | ** | ** | | | ** | | | | | | |
| Cover of areas influenced by man (built-up etc) | ** | | | | * | | | | | | | *** |
| Bedrock geology | | | | | | | | | | | | |
| Amphibolites and greenstones | | * | ** | *** | | ** | | | | | | |
| Gabbros and diorites | | ** | ** | *** | | ** | | | | | | |
| Gneisses | ** | ** | – | *** | | – | | | | | | |
| Chalks and marls | | | | | | | | | | – | | *** |
| Frequency class for limestone | | | | | * | | | | | – | *** | |
| Till | ** | | | ** | | ** | | | | | | |
| Outcropping bedrock | – | | | *** | | ** | | | | | | |
| Gravels and sand | ** | | | – | | ** | | | | | | |
| Eskers | | | | *** | | ** | | | | | | |
| Peat | | | | *** | | ** | | | | | | |
| Bedrock heterogeneity | | | | *** | | | | | | *** | *** | |
| Climate-related variable | | | | | | | | | | | | |
| Annual mean (total) precipitation | ** | | | | | | | | | – | – | |
| Vegetation and floristic variables | | | | | | | | | | | | |
| Cover of forest | ** | ** | ** | | | | | | | | | *** |
| Cover of moist species-rich forests mires | | | | | * | | *** | | | | | |
| Cover of alpine heaths | | | | | | | *** | | | | | |
| Cover of bogs | | ** | ** | | | | | | | | | |
| Cover of pine forests | | | ** | | | | | | | | | |
| Cover of species-rich deciduous forests | | ** | ** | | | | | | | | | |
| Cover of subalpine birch forest | ** | | | | * | | | | | | | |