Journal of Vegetation Science 18: 835-846, 2007
© IAVS; Opulus Press Uppsala.

Semi-dry grasslands along a climatic gradient across Central Europe: Vegetation classification with validation

Illyés, Eszter[1*]; Chytrý, Milan[2,6]; Botta-Dukát, Zoltán[1,7]; Jandt, Ute[3]; Škodová, Iveta[4,8]; Janišová, Monika[4,9]; Willner, Wolfgang[5] & Hájek, Ondřej[2,10]

[1]*Institute of Ecology and Botany, Hungarian Academy of Sciences, H-2163 Vácrátót, Hungary;* [2]*Department of Botany and Zoology, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic;* [3]*Institute of Geobotany and Botanical Garden, Am Kirchtor 1, D-06108 Halle, Germany; E-mail ute.jandt@botanik.uni-halle.de;* [4]*Institute of Botany, Slovak Academy of Sciences, Dúbravská cesta 14, SK-845 23 Bratislava, Slovakia;* [5]*VINCA – Vienna Institute for Nature Conservation and Analyses, Giessergasse 6/7, A-1090 Vienna, Austria; E-mail wolfgang.willner@vinca.at;* [6]*E-mail chytry@sci.muni.cz;* [7]*E-mail bdz@botanika.hu;* [8]*E-mail iveta.skodova@savba.sk; E-mail* [9]*monika.janisova@savba.sk;* [10]*E-mail ohajek@sci.muni.cz;* [*]*Corresponding author; Fax +36 28360110; E-mail illyese@botanika.hu*

Abstract

Question: What is the variation in species composition of Central European semi-dry grasslands? Can we apply a training-and-test validation approach for identifying phytosociological associations which are floristically well defined in a broad geographic comparison; can we separate them from earlier described associations with only a local validity?

Location: A 1200 km long transect running along a gradient of increasing continentality from central Germany via Czech Republic, Slovakia, NE Austria, Hungary to NW Romania.

Methods: Relevés with > 25% cover of *Brachypodium pinnatum* and/or *Bromus erectus* were geographically selected from a larger database. They were randomly split into two data sets, TRAINING and TEST, each with 422 relevés. Cluster analysis was performed for each data set on scores from significant principal coordinates. Different partitions of the TRAINING data set were validated on the TEST data set, using a new method based on the comparison of % frequencies of species occurrence in clusters. Clusters were characterized by statistically defined groups of diagnostic species and values of climatic variables.

Results: Species composition changed along the NW-SE gradient and valid clusters were geographically well separated. Optimal partition level was at 11 clusters, six being valid: two clusters Germany and the Czech Republic corresponded to the *Bromion erecti*; two clusters from the Czech Republic and Hungary to the *Cirsio-Brachypodion*, and two clusters were transitional between these two alliances.

Conclusion: The training-and-test validation method used in this paper proved to be efficient for discriminating between robust clusters, which are appropriate candidates for inclusion in the national or regional syntaxonomic overviews, and weak clusters, which are specific to the particular classification of the given data set.

Keywords: Austria; *Bromion*; *Cirsio-Brachypodion*; Czech Republic; Germany; Hungary; Phytosociology; Romania; Slovakia; Training and test data sets; Vegetation database.

Nomenclature: Ehrendorfer (1973).

Introduction

The past decade has witnessed a rapid development of electronic phytosociological databases (Ewald 2001; Hennekens & Schaminée 2001), which can be used to create vegetation classification schemes valid over large areas and across national boundaries. In Europe, this offers a unique opportunity for international harmonization of vegetation classification, habitat typologies and the subsequent planning of conservation strategies.

However, vegetation units based on numerical classifications of data from selected areas or selected vegetation types are often not appropriate for direct inclusion in large-scale vegetation overviews, because such classifications are highly idiosyncratic. They accurately reflect the structure of the input data set but do not use any external information; therefore some clusters are often specific to the particular classification but are rarely found in the classifications of other data sets from the same vegetation type. National or international systems of vegetation classification, however, should be more robust and include only those vegetation units which have been recognized in several independent classifications. Large-scale vegetation classification projects would be greatly improved if the case studies involving numerical vegetation classification clearly separated clusters with a more general validity from clusters specific to the particular data set. One approach to achieve this is simple validation with a training and test data set (Duda et al. 2001), i.e. making a classification on a training data set, then applying the same classification method to a different (test) data set, comparing the classifications of the training and test data sets, and finally identifying the corresponding clusters (vegetation types) revealed in both data sets. So far, validation has been rarely used in studies describing vegetation patterns across landscapes, mainly due to the limited amounts of available data, which rarely permit the compilation of an independent data set in addition to the data set used for the primary analysis. An exception is the study of Hallgren et al. (1999) who called their approach cross-validation, although they used simple validation with random splitting of the data into training and test data set which is different from cross-validation techniques used in pattern recognition studies (e.g. *m*-fold cross-validation or leave-one-out cross-validation; Duda et al. 2001). The recent emergence of large electronic phytosociological databases provides an opportunity for compiling validation data sets and a wider use of validation or cross-validation procedures in vegetation classification.

The semi-dry grasslands of Central Europe are a suitable model for demonstrating the issues related to vegetation classification at an international level. They are the focus of nature conservation (e.g. the EU Habitats Directive and Natura 2000 network) because of their high species richness and the occurrence of many rare or endangered species (Riecken et al. 1994; Borhidi & Sánta 1999; Chytrý et al. 2001; Stanová & Valachovič 2002). Historically, they were investigated independently in different countries, which resulted in a set of national classifications with only limited international compatibility (Klika 1933; Wagner 1941; Oberdorfer 1993; Krausch 1961; Mahn 1965; Eijsink et al. 1978; Mucina & Kolbek 1993; Borhidi 2003). So far, no comparative analysis has been performed that would establish clear links between corresponding semi-dry grassland types of different countries. Therefore we compiled a geographically stratified data set of phytosociological relevés dominated by *Brachypodium pinnatum* or *Bromus erectus*, characteristic dominants of semi-dry grasslands, from an approximately 1200 km long NW-SE transect, running along the main floristic gradient of these grasslands (Willems 1982) from central Germany to northwestern Romania. In the subatlantic northwest, most of these grasslands probably developed as secondary vegetation after the deforestation of mesic forests (Willems 1982; Mucina & Kolbek 1993), while in the subcontinental southeast, particularly in the Carpathian Basin, many of them may be natural components of the forest-steppe landscapes (Zólyomi & Fekete 1994). This implies that a detailed knowledge of community variation within Central European semi-dry grasslands may provide the scientific basis for designing management plans that would be more suitable for maintaining the biodiversity of particular landscapes.

The aim of this paper is (1) delimitation and description of major vegetation types of Central European semi-dry grasslands and (2) exploration of some issues related to the international standardization of vegetation typologies, in particular testing the newly developed training-and-test validation method with real data.

Material and Methods

Vegetation data

We collected 13 412 relevés of semi-dry grasslands from a geographic and macroclimatic gradient running from central Germany through the Czech Republic, Austria, Slovakia and Hungary to northwestern Romania. The German relevés were from the database compiled by Jandt (1999), the Czech and Slovak relevés from the respective Czech and Slovak national phytosociological databases (Chytrý & Rafajová 2003; Valachovič 1999). The Hungarian relevés were partly collected from the literature and unpublished sources and partly newly recorded by E. Illyés; presently they are stored in the Hungarian national phytosociological database (CoenoDat). The Austrian and Romanian relevés were mostly taken from local literature. We only selected relevés from plots ≥ 4 m^2 and ≤ 100 m^2.

A particular problem was the formal delimitation of the study object: semi-dry grassland vegetation. We could not base our relevé selection on syntaxonomical categories, because classification schemes of these grasslands are rather arbitrary and differ between countries (e.g. Mucina & Kolbek 1993; Oberdorfer 1993; Borhidi 2003; Chytrý 2007). Therefore we only selected relevés in which at least one of the grasses *Brachypodium pinnatum* and *Bromus erectus* occurred with a cover > 25% and which were assigned to the phytosociological class of dry grasslands, *Festuco-Brometea*. This selection yielded 2926 relevés. *Brachypodium pinnatum* and *Bromus erectus* are frequently dominant in Central European semi-dry grasslands, so their dominance could be used as an operational criterion for the inclusion of relevés in our data set. Bryophyte and lichen records were excluded since they were missing in many relevés; generally, cryptogams are not very common in these grasslands. Taxonomically difficult species were merged into aggregated species (e.g. *Brachypodium pinnatum* and *B. rupestre* were merged into *B. pinnatum*). For the analysis we replaced the cover estimates contained in the original data by presences/absences, because our validation method uses this data type for the calculation.

Data stratification, training and test data sets

Large phytosociological data sets compiled from heterogeneous sources often contain many relevés from some small areas where sampling was more intensive than elsewhere. In order to prevent such local oversampling affecting the analysis, we tried to increase the representativeness of our data set by geographically stratified resampling (Knollová et al. 2005). We randomly selected a maximum of five relevés from each cell of a geographic grid of 6' latitude and 10' longitude.

Then we randomly split the resampled data set into two subsets of equal size, hereafter called TRAINING and TEST, with the aim of using the former to create the classification and the latter for validating the individual clusters resulting from this classification. After the split we had 442 relevés in each of the data sets. To remove the undue effect of relevés with outlying species composition, we performed separate outlier analyses for the TRAINING and TEST data sets, using the PC-ORD 4 program (McCune & Mefford 1999) with the Sørensen coefficient. After outlier exclusion, TRAINING and TEST data sets contained 422 relevés each, including 114 and 123 relevés from Germany, 179 and 190 from the Czech Republic, 52 and 49 from Slovakia, 49 and 36 from Hungary, 18 and 15 from Austria and 10 and 9 from Romania, respectively (App. 1).

Cluster analysis

Large vegetation databases may contain a high proportion of noise, i.e. random variation, which can cause artefacts in the numerical classification processes, in particular in agglomerative methods (Gauch 1982a: 208). Noise can be reduced by using the coordinates of the relevés along the ecologically meaningful axes of the principal coordinates analysis (PCoA; Legendre & Legendre 1998) instead of the raw data as input for the classification (Gauch 1982b; Botta-Dukát et al. 2005). We performed PCoA in the R software (www.r-project.org) with the VEGAN package by J. Oksanen (http://cc.oulu.fi/~jarioksa) using presence/absence data with Sørensen dissimilarity (Legendre & Legendre 1998). To determine the PCoA axes that contain interpretable ecological information, we compared the percentage eigenvalues with random expectations based on the broken-stick model (Legendre & Legendre 1998: 410). The number of significant axes was 59 in the TRAINING and 63 in the TEST data set. The significant axes explained 64% and 69% of the total variation in these data sets, respectively. We used the position of the relevés along the significant axes of PCoA as input for the classification. Euclidean distance and Ward's algorithm of minimum increment of sum of squares (Legendre & Legendre 1998) were used for dendrogram construction in the PC-ORD 4 program.

Validation of classifications

The set of relevés used in any analysis is a sample from the statistical population of all possible relevés that satisfy pre-selected criteria defining this population: in our case it was certain plot size, dominance of some species and species composition corresponding to the class *Festuco-Brometea*. Numerical classification methods explore the structure of the sample, but the aim is to explore the structure of the whole statistical population. Some clusters resulting from numerical classification may be artefacts in the sense that they reflect the structure of the sample but not of the statistical population. This means that the same classification method applied to other samples from the same population would not reveal such clusters.

This problem can be overcome by applying the following validation procedure. The set of relevés is randomly split into two subsets of equal size (in our case called TRAINING and TEST) and the same classification procedure is independently applied to each of them (Duda et al. 2001). On each level of the classification hierarchy groups occurring in the corresponding TRAINING and TEST classifications are compared based on the relative frequency of species. The *Z*-statistic (Zar 1999) is used to compare the relative frequencies of each species:

$$Z = \frac{|q_1 - q_2|}{\sqrt{\frac{Q(1-Q)}{n_1} + \frac{Q(1-Q)}{n_2}}} \qquad (1)$$

where q_1 and q_2 are relative frequencies (constancies) in the corresponding groups of TRAINING and TEST data sets, n_1 and n_2 are group sizes, i.e. numbers of relevés assigned to the groups, and

$$Q = \frac{n_1 q_1 + n_2 q_2}{n_1 + n_2} \qquad (2)$$

Z approximately follows the standard normal distribution, thus the corresponding Type I error probability (p) can be calculated easily. Then these p values are combined by the Fisher's omnibus test (Sokal & Rohlf 1995):

$$\chi^2 = -2\sum \ln p_1 \qquad (3)$$

where p_i is the Type I error probability for species i. First χ^2 is calculated for the whole TRAINING and TEST data sets, and then cluster pairs with χ^2 lower than this value are considered as similar (Botta-Dukát 2007). A cluster of the TRAINING data set is regarded to be valid, if there is one and only one similar cluster in the TEST data set. If there is no such cluster in the TEST data set, the cluster is characteristic only for the sample, but not for the whole population. If there is more than one similar cluster in the TEST data set, i.e. the differences between them are arbitrary, the cluster cannot be validated unambiguously.

The number of valid clusters depends on the total number of clusters in the partition. It is low in partitions with few clusters, because the clusters are too large and heterogeneous. As the number of clusters increases, the number of valid clusters also increases, but when the total number of clusters becomes too high, the valid clusters are divided into smaller clusters rather arbitrarily and the number of valid clusters decreases again. This means that we have to search for valid clusters over a wider range of partitions with different numbers of clusters to find the partition with maximum number of valid clusters. In our data sets we tested partitions with 2-12 clusters.

Determination of diagnostic species

The fidelity of species to clusters of the TRAINING data set was calculated in order to determine diagnostic species for each cluster (Chytrý et al. 2002). This calculation was done for the partition that already contained all the valid clusters but at the same time contained the smallest number of non-valid clusters. The fidelity of species to clusters was calculated with presence/absence data, using the phi-coefficient applied to clusters of equalized size (Tichý & Chytrý 2006), as available in the JUICE program (Tichý 2002). The Φ-values are independent of the statistical significance of species occurrence concentration in the relevés of particular clusters, but in the JUICE program, significance can be obtained by a simultaneous calculation of Fisher's exact test. In our case, we considered a species as diagnostic if $phi > 0.3$ and $P < 0.01$. The threshold value of $phi = 0.3$ was selected because it yielded neither too long nor too short lists of diagnostic species for individual clusters. The classification results are summarized in Table 1.

Climatic and geographic data

Climatic data such as mean annual temperature, July temperature, January temperature and mean annual precipitation for relevé locations were derived from the WORLDCLIM database (http://biogeo.berkeley.edu). As the relevés were located on a NW-SE transect, we also defined geographic position as a potential explanatory variable for vegetation patterns. Instead of the usual way of identifying location by simply using longitude and latitude, respectively, we defined a single geographic variable running in the direction of the major gradient in geographic locations. This was defined as the position of relevés on the first PCA axis (CANOCO 4.5 program; ter Braak & Šmilauer 2002) where the longitudes and latitudes of relevés were used as input data. Medians of the climatic variables were calculated for the merged valid clusters of TRAINING and TEST data sets and differences were tested by Kruskal-Wallis non-parametric ANOVA and subsequently by Dunn's *post-hoc* test (Zar 1999).

Results and Discussion

Classification and validation results

The hierarchical level of the dendrogram with the highest number of valid clusters was the one with 11 clusters, of which six were valid. In the TRAINING data set the valid clusters contained altogether 204 relevés (48.3% of the data set), while the remaining 218 relevés (51.2%) belonged to non-valid clusters. In the TEST data set the corresponding figures were 215 (50.9%) and 207 (49.1%). Usually the valid clusters had more diagnostic species than had the non-valid clusters (Table 1) and narrower geographic ranges (App. 2), though some valid clusters had a large range in one of the TRAINING or TEST data sets but a small one in the other.

Dendrogram topographies of the TRAINING and TEST data sets (Fig. 2) reveal that the same pairs of valid clusters form smaller groups (A-B, C-D and E-F) in both data sets. The higher level, i.e. the linkage of the cluster pairs, is different in the two dendrograms, which explains why the higher-level clusters were not confirmed as valid.

Description and interpretation of the classification

There are remarkable differences among the valid clusters in all climatic variables (Table 2). Clusters A and B (subatlantic *Brachypodium* and *Bromus* grasslands) are the most oceanic ones according to geographic position, precipitation and temperature. Clusters C (semi-dry grasslands on wetter soils with wider distribution), D (species-rich meadows, mainly found in the White Carpathians) and E (open subcontinental dry grasslands) have a transitional character, while cluster F (*Brachypodium* grasslands of the inner Carpathian Basin) are confined to the driest and warmest areas.

This pattern shows that species composition of semi-dry grasslands changes considerably along the NW-SE gradient across Central Europe (Willems 1982). In areas characterized by suboceanic climate in central Germany and the middle altitudes of the Czech Republic and Slovakia these grasslands contain subatlantic species such as *Cirsium acaule, Gentianella germanica* agg., *Potentilla neumanniana* and *Thymus pulegioides*. By contrast, in the drier parts of the study area, semi-dry grasslands contain several species of continental distribution, which are also typical of dry oak forests, e.g. *Centaurea triumfettii, Galium glaucum, Geranium sanguineum, Inula ensifolia, I. hirta, Peucedanum cervaria, Tanacetum corymbosum* and *Thesium linophyllon*, or continental steppe species such as *Chamaecytisus austriacus, Linum flavum* and *Stipa capillata*.

Table 1. Synoptic table of 11 clusters of the TRAINING data set with percentage frequency (constancy) of species. Within blocks of diagnostic species, species are ranked by decreasing fidelity, measured by the Φ-coefficient for relevé groups of equalized size (*: *phi* > 0.3; **: *phi* > 0.5). Species with non-significant occurrence concentration in the given cluster were not included in the groups of diagnostic species, even if they had *phi* > 0.3 (Fisher's exact test, *P* < 0.01).

Group	A	B	nv	C	D	nv	nv	nv	E	nv	F
No. of relevés	61	30	49	18	28	71	20	46	25	32	42
A. *Brachypodium* dominated subatlantic grasslands of Central Germany and the Czech Republic											
Koeleria pyramidata agg.	89 *	53	22	28	14	13	15	—	28	19	2
Pinus sylvestris	23 *	—	—	—	—	8	—	—	—	—	—
Hieracium pilosella	61 *	37	16	—	7	4	15	20	28	22	—
Campanula rotundifolia agg.	56 *	10	20	6	4	13	—	17	16	31	17
Melilotus alba	8 *	—	—	—	—	—	—	—	—	—	—
Ranunculus bulbosus	38 *	13	16	6	25	6	5	4	—	6	—
Ononis repens	13 *	3	2	—	—	1	—	—	—	—	—
Anthyllis vulneraria	49 *	33	22	—	29	7	—	22	12	31	—
Linum catharticum	85 *	70	47	61	64	24	20	26	52	69	14
Carex ornithopoda	8 *	3	—	—	—	—	—	—	—	—	—
Polygala chamaebuxus	5 *	—	—	—	—	—	—	—	—	—	—
Avenochloa pratensis	36 *	7	2	—	—	4	25	17	20	34	5
Calluna vulgaris	7 *	—	—	—	—	—	—	2	—	—	—
B. *Bromus* and *Brachypodium* dominated subatlantic grasslands invaded by shrubs											
Medicago lupulina	39	70 *	31	17	25	4	5	4	—	9	—
Gymnadenia conopsea	15	33 *	—	—	14	—	—	—	4	3	2
Rosa canina agg.	25	57 *	27	—	18	24	5	4	4	16	5
Prunus spinosa agg.	10	47 *	16	—	7	14	5	4	8	16	10
Rosa rubiginosa agg.	3	20 *	—	—	—	—	5	4	—	—	—
Gentianella germanica agg.	18	27 *	—	6	—	—	—	—	4	—	—
Frangula alnus	—	17 *	—	6	—	1	—	—	—	—	—
Fraxinus excelsior	5	17 *	—	6	—	—	—	2	—	3	—
Crataegus spec.	8	50 *	29	6	14	17	15	17	8	9	31
Cornus sanguinea	21	33 *	8	17	—	11	5	—	—	9	2
Non-valid group 1											
Polygala vulgaris	3	—	18 *	—	11	1	—	4	—	3	—
Galium pusillum agg.	20	13	22 *	—	—	3	—	—	—	9	—
Corylus avellana	8	—	14 *	—	4	7	—	—	—	—	—
Potentilla recta	—	—	6 *	—	—	1	—	—	—	—	—
C. Semi-dry grasslands on wetter soils dominated by *Bromus erectus*											
Equisetum arvense	—	3	2	33 *	4	3	—	—	—	—	—
Potentilla reptans	2	3	8	44 *	4	10	10	—	—	—	—
Tetragonolobus maritimus	2	—	—	39 *	—	1	—	—	20	3	—
Cirsium tuberosum	—	—	—	17 *	—	—	—	—	—	—	—
Glechoma hederacea agg.	—	3	2	22 *	4	—	5	—	—	—	—
Succisa pratensis	—	—	—	11 *	—	—	—	—	—	—	—
Silaum silaus	—	—	—	11 *	—	—	—	—	—	—	—
Senecio erucifolius	2	7	—	17 *	—	—	—	—	—	—	—
Rubus caesius	—	7	2	17 *	—	1	—	—	—	—	—
Carex hirta	—	—	4	17 *	4	—	—	2	—	—	2
Pastinaca sativa	2	10	2	22 *	—	1	10	—	—	3	—
Agrostis gigantea	2	—	—	17 *	—	3	—	4	4	3	—
D. Mostly *Bromus* dominated grasslands of the White Carpathians											
Campanula patula	—	—	8	6	61 **	3	—	—	—	—	—
Luzula campestris agg.	7	—	20	11	79 **	4	—	11	—	—	5
Cruciata glabra	—	—	27	—	68 **	4	—	—	—	6	2
Anthoxanthum odoratum	11	3	27	6	79 **	6	—	7	—	3	5
Carex pallescens	—	—	—	—	39 **	—	5	—	—	—	—
Rumex acetosa	7	—	12	11	64 **	6	—	4	—	—	17
Trifolium pratense	21	7	16	22	75 **	4	—	4	4	9	2
Trifolium repens	7	—	14	11	50 **	—	—	—	—	3	—
Alchemilla vulgaris agg.	7	—	22	6	50 *	1	—	—	—	—	—
Cerastium holosteoides	7	7	6	11	46 *	—	—	—	—	—	—
Ajuga reptans	—	—	—	—	25 *	—	—	—	—	—	—
Primula veris	21	3	18	22	75 *	8	15	2	8	12	5
Viola canina	2	—	8	—	32 *	—	—	—	—	—	—
Crepis biennis	5	—	4	6	36 *	1	—	—	—	—	—
Prunus domestica	—	3	2	—	29 *	1	—	—	—	—	—
Danthonia decumbens	2	—	10	—	36 *	—	—	2	4	—	—
Leucanthemum vulgare agg.	23	50	51	33	89 *	7	—	11	8	16	7
Carex montana	10	—	14	—	46 *	3	—	—	4	3	10

Group	A	B	nv	C	D	nv	nv	nv	E	nv	F
No. of relevés	61	30	49	18	28	71	20	46	25	32	42
Festuca pratensis agg.	11	—	10	—	54 *	1	20	2	—	3	17
Centaurea phrygia agg.	—	—	4	—	25 *	1	—	—	—	—	—
Trisetum flavescens	13	27	18	6	68 *	14	35	—	—	9	—
Cynosurus cristatus	2	—	4	6	29 *	—	—	—	—	—	—
Vicia cracca agg.	16	3	35	28	71 *	17	15	4	—	3	17
Trifolium montanum	3	—	29	6	79 *	10	10	22	12	41	43
Veronica officinalis	—	—	12	—	29 *	1	—	—	—	—	—
Hypericum maculatum	—	—	4	—	21 *	—	—	—	—	—	—
Stellaria graminea	—	—	2	—	21 *	—	—	2	—	—	—
Colchicum autumnale	—	—	12	6	32 *	1	—	—	—	3	2
Primula vulgaris	—	—	—	—	14 *	—	—	—	—	—	—
Tragopogon pratensis agg.	11	13	4	17	54 *	8	5	11	4	22	5
Aquilegia vulgaris	—	—	4	6	21 *	—	—	—	—	—	—
Cirsium pannonicum	—	—	8	—	39 *	4	—	—	4	22	17
Rhinanthus minor	8	—	14	6	32 *	—	—	4	—	3	—
Festuca rubra agg.	15	7	33	39	57 *	7	15	7	4	3	—
Carex panicea	—	—	2	—	14 *	—	—	—	—	—	—
Hypochoeris maculata	—	—	2	—	25 *	—	—	2	—	9	7
Listera ovata	—	—	8	6	21 *	—	—	—	—	—	—
Potentilla erecta	8	—	16	6	29 *	1	—	—	—	—	—
Plantago lanceolata	61	40	39	67	86 *	15	5	35	4	53	2
Arrhenatherum elatius	18	33	49	28	86 *	48	45	13	8	34	50
Potentilla collina agg.	—	—	—	—	11 *	—	—	—	—	—	—
Prunella vulgaris	26	10	16	33	46 *	3	5	2	—	3	2
Holcus lanatus	7	—	8	6	21 *	1	—	—	—	—	—
Alchemilla glaucescens	—	—	2	—	11 *	—	—	—	—	—	—
Ranunculus auricomus agg.	—	—	2	—	11 *	—	—	—	—	—	—
Carum carvi	—	—	4	11	18 *	—	—	—	—	—	—
Lathyrus latifolius	—	—	—	—	21 *	3	—	—	—	9	12
Arabis hirsuta agg.	—	3	8	6	29 *	3	—	9	—	16	2
Trifolium medium	10	13	31	6	39 *	8	15	—	4	6	2
Carpinus betulus	8	—	14	6	25 *	8	—	—	—	—	2
Orchis morio	—	—	—	—	7 *	—	—	—	—	—	—
Hypochoeris radicata	—	—	—	—	7 *	—	—	—	—	—	—
Dactylorhiza sambucina	—	—	—	—	7 *	—	—	—	—	—	—
Lychnis flos-cuculi	—	—	—	—	7 *	—	—	—	—	—	—
Prunella laciniata	—	—	10	6	21 *	—	—	2	—	9	2
Phyteuma spicatum	2	—	2	—	11 *	1	—	—	—	—	—
Allium scorodoprasum	—	—	2	6	14 *	6	—	—	—	—	—
Crepis praemorsa	—	—	—	—	11 *	—	—	—	8	—	—
Myosotis arvensis	—	3	4	6	14 *	—	5	—	—	—	—
Briza media	75	73	65	78	86 *	10	20	35	44	62	38
Avenochloa pubescens	11	—	16	—	32 *	10	5	4	4	12	31
Non-valid group 2											
Coronilla varia	18	3	45	33	29	63 *	20	30	12	56	29
Origanum vulgare	5	7	24	—	7	27 *	—	—	4	9	10
Non-valid group 3											
Galium verum agg.	43	10	57	83	54	49	100 *	63	24	56	50
Cirsium eriophorum	—	—	—	—	—	3	15 *	2	—	—	2
Poa pratensis agg.	44	37	65	33	46	59	80 *	35	12	31	45
Non-valid group 4											
Centaurea paniculata agg.	3	—	2	—	—	3	—	33 *	4	6	—
Astragalus onobrychis	—	—	—	—	—	—	—	30 *	4	16	10
Eryngium campestre	2	—	2	11	—	20	10	59 *	40	31	21
Nonea pulla	—	—	—	—	—	1	—	15 *	—	—	2
Salvia nutans	—	—	—	—	—	—	—	9 *	—	—	—
Artemisia campestris	—	—	—	—	—	—	—	11 *	—	—	2
Festuca valesiaca	—	—	4	—	—	4	5	24 *	—	16	5
Campanula sibirica	—	—	—	—	—	1	—	17 *	4	9	2
Allium flavum	—	—	—	—	—	—	—	7 *	—	—	—
Silene otites	—	—	—	—	—	—	—	9 *	—	—	2
Stipa capillata	—	—	—	—	—	1	5	15 *	8	—	5
Euphorbia seguieriana	—	—	—	—	—	1	—	7 *	—	—	—
Peucedanum oreoselinum	—	—	4	—	—	1	—	11 *	4	—	—
Veronica spicata agg.	—	—	8	—	4	—	—	22 *	4	19	10
Seseli pallasii	—	—	—	—	—	—	—	7 *	—	—	2

Group	A	B	nv	C	D	nv	nv	nv	E	nv	F
No. of relevés	61	30	49	18	28	71	20	46	25	32	42
E. Open grasslands on calcareous bedrock mostly from Bohemia											
Thymus praecox	13	—	—	11	—	4	—	13	68 **	6	—
Linum tenuifolium	—	—	—	—	—	—	—	13	32 *	19	2
Biscutella laevigata	—	—	—	—	—	—	—	—	8 *	—	—
Bromus pannonicus	—	—	—	—	—	—	—	—	8 *	—	—
Coronilla vaginalis	—	—	—	—	—	—	—	—	8 *	—	—
Jurinea mollis	—	—	—	—	—	—	—	—	8 *	—	—
Globularia punctata	—	—	2	—	—	1	—	2	16 *	6	2
Helianthemum canum	—	—	—	—	—	1	—	2	12 *	3	—
Euphorbia cyparissias	54	40	61	11	25	70	55	65	84 *	62	19
Non-valid group 5											
Chamaecytisus ratisbonensis	—	—	—	—	—	1	—	9	4	59 **	10
Aster amellus	—	3	—	—	—	10	—	—	12	66 **	19
Polygala major	—	—	8	—	7	1	—	4	8	47 *	12
Scabiosa ochroleuca	3	7	14	17	4	28	—	46	48	84 *	19
Stachys recta	7	—	2	—	—	20	—	20	4	47 *	29
Seseli libanotis	—	—	—	—	—	4	—	2	—	19 *	—
Viola rupestris	2	—	2	—	—	—	—	4	4	22 *	—
Pulsatilla grandis	—	—	—	—	—	4	—	9	4	28 *	12
Buphthalmum salicifolium	—	—	—	—	7	4	—	—	—	19 *	—
Prunella grandiflora	16	3	8	6	—	7	5	4	20	41 *	10
Hypericum elegans	—	—	—	—	—	—	—	—	—	9 *	—
Salvia pratensis	15	13	20	39	75	56	40	48	60	94 *	67
Orchis militaris	2	—	2	—	—	—	—	—	—	12 *	—
Thymus pannonicus	—	—	2	—	—	3	5	20	—	25 *	5
Anthericum ramosum	2	3	2	—	7	13	—	17	20	38 *	31
Onobrychis viciifolia agg.	13	3	8	—	14	10	—	17	24	41 *	21
Orobanche gracilis	—	—	—	—	—	—	—	—	—	6 *	—
Thymus glabrescens	—	—	4	—	—	6	—	15	—	22 *	14
Peucedanum cervaria	5	—	6	—	4	7	5	7	16	28 *	24
Euphorbia virgata	—	—	—	6	7	1	—	9	—	19 *	12
F. *Brachypodium* grasslands of the inner Carpathian Basin											
Euphorbia pannonica	—	—	—	—	—	1	—	2	8	—	62 **
Avenochloa adsurgens	—	—	—	—	—	—	—	2	—	—	38 **
Chamaecytisus austriacus	—	—	—	—	—	—	—	7	—	9	45 **
Agropyron intermedium	—	—	—	—	—	3	15	11	4	12	55 **
Tanacetum corymbosum agg.	—	—	14	—	4	10	5	9	4	25	62 *
Linum flavum	—	—	—	—	—	—	—	—	—	6	26 *
Thalictrum minus	—	—	—	—	—	4	—	7	—	9	33 *
Hieracium umbellatum	—	—	2	—	—	3	—	—	—	9	29 *
Galium glaucum	—	—	—	—	—	7	5	13	4	6	38 *
Peucedanum alsaticum	—	—	—	—	—	3	—	4	—	16	31 *
Lathyrus pannonicus	—	—	—	—	—	—	—	—	—	—	14 *
Inula hirta	—	3	2	—	—	6	5	2	4	6	29 *
Campanula bononiensis	—	—	—	—	—	—	5	—	—	—	14 *
Thesium arvense	—	—	—	—	—	—	—	4	—	—	12 *
Trifolium alpestre	—	—	18	—	4	4	5	9	8	19	33 *
Phleum phleoides	7	—	4	—	4	10	—	26	—	28	36 *
Veronica austriaca	—	—	—	—	—	—	—	2	4	9	17 *
Medicago prostrata	—	—	—	—	—	—	—	—	—	—	7 *
Adonis vernalis	2	—	2	—	—	4	10	15	8	3	26 *
Aster linosyris	—	—	—	—	—	1	—	11	16	22	26 *
Verbascum lychnitis	—	—	—	—	—	4	—	4	—	—	12 *
Pulmonaria mollis	—	—	2	—	4	—	—	—	—	3	12 *
Prunus fruticosa	—	—	—	—	—	1	—	—	—	—	7 *
Serratula radiata	—	—	—	—	—	—	—	—	—	—	5 *
Myosotis ramosissima	—	—	—	—	—	—	—	—	—	—	5 *
Torilis arvensis	—	—	—	—	—	—	—	—	—	—	5 *
Species diagnostic for more than one cluster											
Potentilla neumanniana	64 *	60 *	14	—	—	7	5	4	8	—	—
Scabiosa columbaria	61 *	60 *	8	22	—	—	5	2	4	—	—
Festuca ovina	62 *	77 **	16	6	—	6	15	—	—	—	—
Carex flacca	69 *	53 *	24	44	14	10	—	—	20	22	—
Sanguisorba minor	98 *	83	57	22	50	55	30	33	92 *	56	7
Thymus pulegioides	64 *	70 *	47	11	43	10	15	4	4	16	—
Leontodon hispidus	72 *	43	61	39	86 *	8	5	2	20	47	26
Carex caryophyllea	46 *	20	20	—	46 *	4	5	15	—	38	7
Lotus corniculatus agg.	79	90 *	63	17	93 *	37	15	35	64	66	12
Veronica chamaedrys agg.	7	—	47 *	6	75 **	17	—	4	—	3	7

Group	A	B	nv	C	D	nv	nv	nv	E	nv	F
No. of relevés	61	30	49	18	28	71	20	46	25	32	42
Agrostis tenuis	11	—	43 *	17	50 *	4	25	11	12	—	—
Ranunculus acris	—	—	8	28 *	32 *	1	—	—	—	—	—
Taraxacum officinale agg.	25	3	24	39 *	54 *	8	—	4	—	3	5
Campanula glomerata	8	—	4	—	57 *	4	5	2	8	56 *	40
Filipendula vulgaris	3	—	16	6	75 *	17	55	37	16	19	74 *
Dactylis glomerata	34	23	37	50	96 *	35	80 *	24	16	66	67
Betonica officinalis	—	—	24	—	36 *	4	5	2	—	3	38 *
Carlina acaulis	16	10	45	—	50 *	15	5	7	12	66 *	2
Koeleria macrantha	3	—	6	—	—	7	10	59 *	16	53 *	19
Carex humilis	—	—	2	—	4	8	10	63 *	68 *	38	21
Potentilla arenaria	3	3	2	—	4	3	5	46 *	24	38 *	—
Astragalus austriacus	—	—	—	—	—	—	—	20 *	16 *	6	—
Bothriochloa ischaemum	—	—	—	—	—	3	—	20 *	—	25 *	—
Seseli hippomarathrum	—	—	—	—	—	3	5	22 *	20 *	9	—
Dianthus carthusianorum agg.	2	—	14	—	36	18	10	43 *	4	44 *	19
Asperula cynanchica	3	3	16	6	7	20	5	30	72 *	78 *	26
Bupleurum falcatum	10	10	14	—	—	35	10	20	64 *	72 *	10
Scabiosa canescens	—	—	—	—	—	3	15	17	28 *	31 *	2
Thesium linophyllon	—	—	4	6	7	7	—	15	8	56 *	36 *
Centaurea scabiosa agg.	39	27	29	17	29	45	10	43	52	97 *	74 *
Festuca rupicola	7	7	49	22	57	68	35	80	40	100 *	88 *
Dorycnium pentaphyllum agg.	—	—	12	—	4	7	—	30	16	47 *	36 *
Carex michelii	—	—	2	—	4	6	—	—	4	34 *	43 *
Seseli annuum	2	—	—	—	—	3	—	20	4	34 *	40 *
Inula ensifolia	—	—	2	—	—	1	—	7	12	28 *	26 *
Cirsium acaule	74 *	77 *	29	22	—	8	35	17	60 *	—	—
Bromus erectus	20	83 *	20	94 *	89 *	37	15	54	44	56	24
Ranunculus polyanthemos agg.	2	3	16	17	50 *	8	—	9	8	44 *	43 *
Teucrium chamaedrys	2	3	22	—	7	28	15	43	56 *	56 *	62 *
Other species with frequency > 20%											
Brachypodium pinnatum	97	70	94	56	46	83	100	70	88	100	98
Medicago falcata	15	—	33	6	39	35	50	50	36	56	52
Polygala comosa	39	37	27	17	39	6	—	4	8	—	5
Helianthemum nummularium agg.	30	—	29	—	46	25	5	20	28	34	12
Ononis spinosa	38	10	37	33	21	4	—	22	44	38	5
Agrimonia eupatoria	39	37	57	28	21	38	45	13	16	12	40
Salvia verticillata	3	—	29	6	25	18	—	9	12	19	5
Galium mollugo agg.	34	43	41	28	39	35	10	2	—	—	5
Centaurea jacea	44	17	43	33	39	23	35	15	56	28	24
Viola hirta	43	67	49	28	68	34	15	17	24	47	29
Hieracium bauhinii	3	—	18	—	25	3	—	9	16	12	5
Fragaria vesca	10	20	24	—	14	6	5	2	4	—	—
Daucus carota	34	30	39	28	32	7	10	9	—	25	—
Pimpinella saxifraga agg.	67	47	73	56	82	54	35	48	60	75	43
Knautia arvensis agg.	59	20	57	11	50	55	15	24	36	66	45
Campanula rapunculoides	11	7	16	6	21	8	5	—	12	3	2
Inula salicina	2	—	4	17	18	6	—	4	20	6	7
Hypericum perforatum	39	53	49	22	32	35	30	26	12	50	2
Picris hieracioides agg.	13	—	12	28	—	8	—	2	12	19	5
Achillea millefolium agg.	51	23	90	61	93	77	45	65	48	72	60
Carlina vulgaris agg.	56	57	24	22	4	10	15	13	44	56	14
Plantago media agg.	69	30	67	56	71	41	5	54	52	75	45
Fragaria viridis	23	17	39	6	32	56	45	43	8	56	33
Carex tomentosa	2	—	22	6	18	4	—	2	12	6	2
Potentilla heptaphylla agg.	8	3	33	28	29	25	5	9	32	28	5
Senecio jacobaea	13	20	18	11	7	6	—	11	4	9	—

This provides the basis for the traditional phytosociological division of the alliances *Bromion erecti* (subatlantic group) and *Cirsio-Brachypodion pinnati* (subcontinental group) (Krausch 1961; Mahn 1965; Royer 1991; Mucina & Kolbek 1993; Chytrý 2007). Our classification seems to confirm this separation, with clusters A and B belonging to the former and E and F to the latter alliance. Clusters C and D represent transitional vegetation types between these two alliances, C being confined to specific habitats (wetter soils) and D representing a locally specific vegetation type.

The artificially defined 25% cover limit of *Brachypodium* or *Bromus* in the relevés selected for this analysis makes it impossible to interpret our valid clusters directly in terms of the traditional phytosociological syntaxa, because syntaxa also include stands with similar species composition but lower cover of these grasses. Still, when compared with the Central European phytosociological literature, the valid clusters can be linked to the traditional associations. The species composition, geographic range, climatic features and syntaxonomy of the valid clusters can be summarized as follows:

Clusters A and B: These grasslands, mostly dominated by *Brachypodium pinnatum*, are found in areas with relatively cool summers and high precipitation, especially in central Germany and the submontane areas of the western Czech Republic (Table 2, Fig. 1 and App. 2). The diagnostic species, e.g. *Anthyllis vulneraria, Carex flacca, Linum catharticum, Potentilla neumanniana*, *Ranunculus bulbosus* and *Scabiosa columbaria* are indicators of calcareous soils, which are usually medium-deep rendzinas or pararendzinas over limestone or other calcareous bedrocks. At the same time, the occurrence of species adapted to low-pH soils (e.g. *Festuca ovina* and *Calluna vulgaris*) indicates leaching of carbonates, typical of areas with higher rainfall. These grasslands are of secondary origin, developed after the clearing of *Fagus* or *Quercus-Carpinus* forests and subsequent grazing by sheep and/or goats (Oberdorfer 1993). Cluster A represents managed or recently abandoned stands, while Cluster B represents successional stages after abandonment, as indicated by the occurrence of shrubs, e.g. *Crataegus* spp., *Cornus sanguinea*, *Rosa* spp. and *Prunus spinosa*. This vegetation corresponds to the association *Carlino acaulis-Brometum erecti* Oberdorfer 1957, which is also frequently called *Gentiano-Koelerietum pyramidatae* Knapp ex Bornkamm 1960.

Cluster C: These semi-dry grasslands are usually found on the footslopes, often in a contact zone between semi-dry grasslands and intermittently wet *Molinion* meadows. The specific topographic position and the good water-holding capacity of soils make such habitats wetter than other types of *Brachypodium* and *Bromus* grasslands, but the areas of distribution of this vegetation are macroclimatically rather dry (Table 2). The dominant species is usually *Bromus erectus* and diagnostic species are indicators of mesic or intermittently wet soils (*Equisetum arvense*, *Glechoma hederacea, Potentilla reptans*, *Pastinaca sativa* and *Ranunculus acris*). This cluster has a broad geographic range (Fig. 1 and App. 2) from central Germany through the Czech Republic and Slovakia to southern Hungary. This vegetation has been traditionally assigned to several associations, within which it was often considered as a transitional type to other associations. Studnička (1980) described this vegetation as the *Potentillo reptantis-Caricetum flaccae* association. Although this type is well delimited in the current data set, it tends to be neglected in the local phytosociological literature.

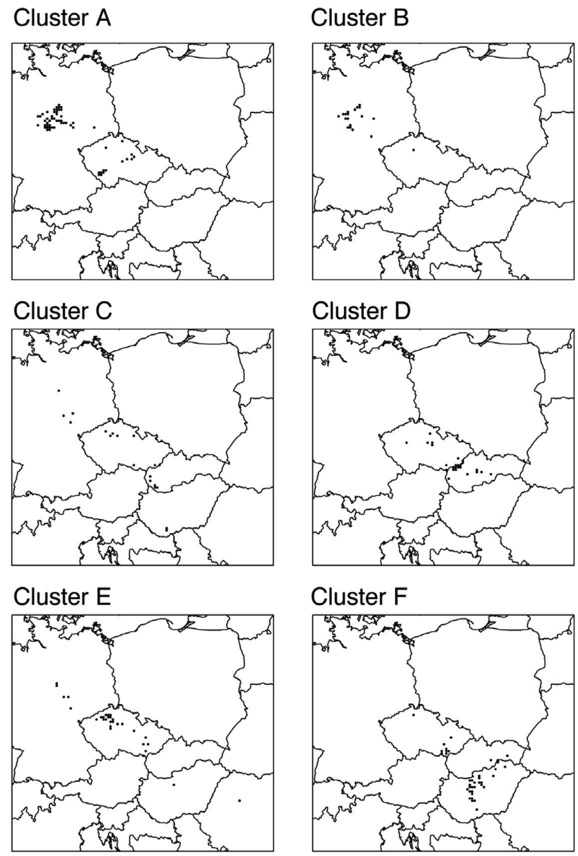


Fig. 1. Distribution maps of relevés of validated clusters A-F, based on the pooled data from the TRAINING and TEST data sets.

Cluster D: Most relevés of this cluster are from the White Carpathians, a mountain range on the border between the Czech Republic and Slovakia. This area is very close to the dry areas with Pannonian steppic flora in the southeastern Czech Republic (southern Moravia) and western Slovakia, but at the same time it receives higher precipitation (650-850 mm/year) than other dry grasslands of Central Europe (Table 2). Some sites from other parts of the Czech Republic and Slovakia also belong to this cluster (Fig. 1). The relevés in our data sets are dominated by *Bromus* or *Brachypodium*, but grasslands of similar species composition can also be dominated by *Molinia arundinacea* or *Carex montana*. These grasslands combine species of mesic meadows, steppes and oligotrophic submontane grasslands. If regularly cut, they contain 60-80 species per 16-25 m^2, thus belonging to the most species-rich grasslands of temperate Europe (Klimeš 1997). They occur on gentle

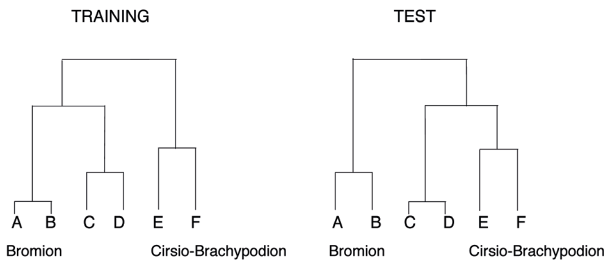


Fig. 2. Topography of dendrograms based on the TRAINING and TEST data sets. Only valid clusters are shown.

slopes with deep soils over calcareous flysch sandstones and claystones. Outcrops of water-holding claystones may cause local waterlogging, but in dry periods of the year these places dry out, which supports species adapted to intermittently wet soils, such as *Betonica officinalis* and *Filipendula vulgaris*. The topsoil is usually slightly decalcified but a higher pH is maintained below (Tlusták 1975). The origin of these grasslands is secondary: they originated after the clearing of *Fagus*, *Quercus-Carpinus* and *Quercus* forests. They largely correspond to *Brachypodio pinnati-Molinietum arundinaceae* Klika 1939, and partly also to other species-rich grasslands which are transitional between the class *Festuco-Brometea* and the mesic meadows of the alliance *Arrhenatherion*.

Cluster E: These are open grasslands of steep slopes on calcareous bedrocks, occurring mostly in continental areas in Bohemia, but also in Moravia and Germany (Fig. 1). Isolated sites are found in Hungary and Romania. The climate is subcontinental, with rather low annual precipitation and hot summers (Table 2). The stands are dominated by *Brachypodium* or *Bromus*, although in some sites, narrow-leaved caespitose graminoids such as *Carex humilis* and *Festuca rupicola* can also be prominent. In the driest areas, they are usually found on north-facing slopes or footslopes, often in contact with narrow-leaved *Stipa-Festuca* dry grasslands. In areas with higher precipitation, they occupy the driest south-facing slopes. These grasslands are mostly secondary, developed as a replacement vegetation for oak, hornbeam or beech forests, but in some places they may be natural grasslands preserved for millennia on steep south-facing slopes, especially on slopes affected by solifluction and landslides (Studnička 1980). This vegetation corresponds to the *Scabioso ochroleucae-Brachypodietum pinnati* Klika 1933, but in different countries, these grasslands were traditionally assigned to different, locally described associations, e.g. in Germany to the *Adonido-Brachypodietum* (Libbert 1933) Krausch 1961, *Scorzonero hispanicae-Brachypodietum* Gauckler 1957 or *Festuco rupicolae-Brachypodietum* Mahn 1965, and in Slovakia to the *Salvio verticillatae-Brachypodietum* Ružičková 1986.

Cluster F: These are closed, dense and species-rich *Brachypodium* grasslands from the Pannonian region (Fig. 1, App. 2). They are most common in the loess area of Mezőföld in central Hungary and in northern Hungary, southern Slovakia and southern Moravia. The climate is continental: the mean annual and July temperature and the January-July temperature difference is the highest of all clusters, while precipitation is low (Table 2). They are typical of calcareous soils, developed mainly on deeper loess or other Quaternary and Tertiary sediments. These grasslands are very rich in species, have a relatively high proportion of forest-steppe, forest-fringe and dry oak forest species (Fekete et al. 1998), and are usually dominated by *Brachypodium pinnatum*. They have a well-developed vertical structure and contain many broad-leaved herbs and tall forbs (Varga et al. 2000). The present stands are partly considered to be of primary origin, predominantly on extremely steep slopes, but mostly they are regarded as the extended and stabilized clearings of former forest-steppe forests (Borhidi 2003, Varga et al 2000). This type corresponds to the *Polygalo majoris-Brachypodietum pinnati* Wagner 1941 or *Verbasco austriaci-Inuletum ensifoliae* Tlusták 1975. In Hungary these grasslands were assigned to the broadly delimited association *Salvio nemorosae-Festucetum rupicolae* Zólyomi ex Soó 1964 (Borhidi 2003), but Horváth (2002) recently proposed separating them into a narrower association, *Euphorbio pannonicae-Brachypodietum pinnati*.

Table 2. Comparison of the geographic position (relative scores on the NW-SE axis) and climatic variables for the valid clusters of the TRAINING and TEST data sets. Values are medians. Clusters in columns with the same letter do not differ significantly (Dunn's test; $P < 0.05$).

	cluster A	cluster B	cluster C	cluster D	cluster E	cluster F
Geographic position NW-SE	–1.19 ^{a}	–1.33 ^{a}	0.14 ^{bc}	0.78 ^{cd}	–0.17 ^{b}	1.22 ^{d}
Mean January temperature (°C)	–0.9 ^{c}	–0.6 ^{d}	–1.5 ^{bc}	–3.3 ^{a}	–2.2 ^{b}	–1.6 ^{b}
Mean July temperature (°C)	16.5 ^{a}	16.4 ^{a}	18.3 ^{cd}	17.4 ^{b}	17.6 ^{bc}	20.7 ^{d}
Mean annual temperature (°C)	8.0 ^{a}	8.1 ^{a}	8.8 ^{b}	7.8 ^{a}	7.1 ^{a}	10.5 ^{b}
Difference between Jan-Jul temperature (°C)	17.30 ^{b}	17.00 ^{a}	20.15 ^{cd}	20.60 ^{d}	19.75 ^{c}	22.30 ^{e}
Precipitation (mm/year)	719 ^{b}	742 ^{b}	569 ^{a}	723 ^{b}	537 ^{a}	560 ^{a}

Valid and non-valid clusters, robust and vague syntaxa

The training-and-test validation method used in the present study is one possibility for the critical interpretation of clusters resulting from numerical classification. The fact that about half of the relevés of both the TRAINING an TEST data sets belonged to clusters which were not identified by the same analysis of a very similar data set clearly demonstrates that results from the numerical analyses, even those based on large data sets, should be interpreted with caution. Classifications based on numerical procedures may contain both robust clusters, which will be frequently recovered by other analyses in other data sets, and weak clusters, which are specific to the given classification of the given data set. Training-and-test validation seems to be a promising approach to discriminate between robust clusters, i.e. good candidates for obtaining the status of a formal syntaxon and being included in syntaxonomic overviews, and weak clusters with limited validity.

Most of the valid clusters in our analysis had smaller geographic ranges and more diagnostic species than the non-valid clusters (Table 1, App. 2). This suggests that Central European semi-dry grasslands consist of a few geographically restricted types with ecologically specialized species, and other types, which mainly contain generalist species and have rather uniform species composition across large areas. Syntaxa are usually defined so as to include vegetation stands rich in specialized species, while the stands composed mainly of generalist species are often not considered in syntaxonomical systems, even if they cover large areas in landscapes (Kopecký & Hejný 1978). Some attempts were made to include vegetation types without specialist species into the syntaxonomical systems by giving them a separate status of basal or derivative communities (Kopecký & Hejný 1978) or central syntaxa (Dierschke 1981). Our trial with the training-and-test validation of numerical classification suggests that such vegetation types are hardly robust due to the absence of specialist species, i.e. due to the lack of discrimination criteria against other vegetation types. If such vegetation types are included in syntaxonomic systems, they should preferably be broadly delimited, while locally restricted syntaxa that lack specialist species should better be avoided.

Acknowledgements. K. Baráth, A. Csecserics, J. Csiky, M. Duchoslav, J. Erős-Honti, A. Horváth, J. Kállayné Szerényi, A. Kun, A. Lengyel, A. Mesterházi, Cs. Molnár, Zs. Molnár, J. Novák, D. Purger, T. Rédei, E. Ruprecht, G. Schmidt, I. Somodi, G. Szollát, I.J. Türke, E. Uhliarová, K. Ujházy and K. Virágh gave us their unpublished relevés for the analysis, A. Szabó collected the Romanian literature, J. Oksanen helped us with the R program, L. Tichý with the JUICE program, and Philip Dixon and two anonymous referees provided helpful comments on the previous version of the manuscript. This study was funded through the Czech-Hungarian academic exchange programme; M. C. and O. H. also received support from the projects MSM 0021622416 and GA ČR 206/05/0020; I. Š. and M. J. were suported from the projects APVT-51-015804 and VEGA 2/5084/25.

References

Borhidi, A. 2003. *Magyarország növénytársulásai*. Akadémiai Kiadó, Budapest, HU.

Botta-Dukát, Z. 2007. *Validation of hierarchical classifications*. Institute of Ecology and Botany, HAS, Vácrátót, HU (www.botanika.hu/bdz/crossvalid.pdf).

Botta-Dukát, Z., Chytrý, M., Hájková, P. & Havlová, M. 2005. Vegetation of lowland wet meadows along a climatic continentality gradient in Central Europe. *Preslia* 77: 89-111.

Chytrý, M. (ed.) 2007. *Vegetace České republiky 1. Travinná a keříčková vegetace*. Academia, Praha, CZ.

Chytrý, M. & Rafajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75: 1-15.

Chytrý, M., Kučera, T. & Kočí, M. (eds.) 2001. *Katalog biotopů České republiky.* Agentura ochrany přírody a krajiny ČR, Praha, CZ.

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

Dierschke, H. 1981. Zur syntaxonomischen Bewertung schwach gekennzeichneter Pflanzengesellschaften. In: Dierschke, H. (ed.) *Syntaxonomie,* pp. 109-122. J. Cramer, Vaduz, FL.

Duda, R.O., Hart, P.E. & Stork, D.G. 2001. *Pattern classification*. 2nd. ed. John Wiley & Sons, New York, NY, US.

Ehrendorfer, F. (ed.) 1973. *Liste der Gefäßpflanzen Mitteleuropas*. 2nd. ed. Gustav Fischer Verlag, Stuttgart, DE.

Eijsink, J., Ellenbroek, G., Holzner, W. & Werger, M.J.A. 1978. Dry and semi-dry grasslands in the Weinviertel, Lower Austria. *Vegetatio* 36: 129-148.

Ewald, J. 2001. Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. *Ber. R. Tüxen Ges*. 13: 53-69.

Fekete, G., Virágh, K., Aszalós, R. & Orlóci, L. 1998. Landscape and coenological differentiation of *Brachypodium pinnatum* grasslands in Hungary. *Coenoses* 13: 39-53.

Gauch, H.G., Jr. 1982a. *Multivariate analysis and community structure*. Cambridge University Press, Cambridge, UK.

Gauch, H.G., Jr. 1982b. Noise reduction by eigenvector ordinations. *Ecology* 63: 1643-1649.

Hallgren, E., Palmer, M.W. & Milberg, P. 1999. Data diving with cross-validation: an investigation of broad-scale gradients in Swedish weed communities. *J. Ecol.* 87: 1037-1051.

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

Horváth, A. 2002. *A mezőföldi löszvegetáció términtázati szerveződése*. Scientia Kiadó, Budapest, HU.

Jandt, U. 1999. Kalkmagerrasen am Südharzrand und im Kyffhäuser. Gliederung im überregionalen Kontext, Verbreitung, Standortsverhältnisse und Flora. *Diss. Bot.* 322: 1-246.

Klika, J. 1933. Studien über die xerotherme Vegetation Mitteleuropas. II. Xerotherme Gesellschaften in Böhmen. *Beih. Bot. Centralbl., Abt. II,* 50: 707-773.

Klika, J. 1939. Die Gesellschaften des *Festucion vallesiacae*-Verbandes in Mitteleuropa. *Stud. Bot. Čech.* 2/3: 117-157.

Klimeš, L. 1997. Druhové bohatství luk v Bílých Karpatech. *Sborn. Přír. Klubu Uh. Hradiště* 2: 31-42.

Knollová, I., Chytrý, M., Tichý, L. & Hájek, O. 2005. Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *J. Veg. Sci.* 16: 479-486.

Kopecký, K. & Hejný, S. 1978. Die Anwendung einer 'deduktiven Methode syntaxonomischer Klassifikation' bei der Bearbeitung der strassenbegleitenden Pflanzengesellschaften Nordostböhmens. *Vegetatio* 36: 43-51.

Krausch, H.D. 1961. Die kontinentalen Steppenrasen (*Festucetalia valesiacae*) in Brandenburg. *Feddes Repert. Beih.* 139: 167-227.

Legendre, P. & Legendre, L. 1998. *Numerical ecology.* 2nd. ed. Elsevier, Amsterdam, NL.

Mahn, E.-G. 1965. Vegetationsaufbau und Standortsverhältnisse der kontinental beeinflußten Xerothermrasengesellschaften Mitteldeutschlands. *Abh. Sächs. Akad. Wiss., Math.-Naturw. Kl.* 49: 1-138.

McCune, B. & Mefford, M.J. 1999. *PC-ORD. Multivariate analysis of ecological data. Version 4.* MjM Software Design, Gleneden Beach, OR, US.

Mucina, L. & Kolbek, J. 1993. *Festuco-Brometea*. In: Mucina, L., Grabherr, G. & Ellmauer, T. (eds.) *Die Pflanzengesellschaften Österreichs. Teil I. Anthropogene Vegetation*, pp. 420-481. Gustav Fischer Verlag, Jena, DE.

Oberdorfer, E. (ed.) 1993. *Süddeutsche Pflanzengesellschaften. Teil II: Sand- und Trockenrasen, Heide- und Borstgras-Gesellschaften, alpine Magerrasen, Saum-Gesellschaften, Schlag- und Hochstauden-Fluren.* 3rd. ed. Gustav Fischer Verlag, Jena, DE.

Riecken, U., Ries, U. & Ssymank, A. 1994. *Rote Liste der gefährdeten Biotoptypen der Bundesrepublik Deutschland.* Kilda Verlag, Greven, DE.

Royer, J.M. 1991. Synthèse eurosibérienne, phytosociologique et phytogéographique de la classe des *Festuco-Brometea*. *Diss. Bot.* 178: 1-296.

Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*. 3rd. ed. W.H. Freeman & Co., San Francisco, CA, US.

Stanová, V. & Valachovič, M. (eds.) 2002. *Katalóg biotopov Slovenska*. DAPHNE, Bratislava, SK.

Studnička, M. 1980. Vegetace bílých strání Českého středohoří a dolního Poohří. *Preslia* 52: 155–176.

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, NY, US.

Tichý, L. 2002. JUICE, software for vegetation classification. *J. Veg. Sci.* 13: 451-453.

Tichý, L. & Chytrý, M. 2006. Statistical determination of diagnostic species for site groups of unequal size. *J. Veg. Sci.* 17: 809-818.

Tlusták, V. 1975. *Syntaxonomický přehled travinných společenstev Bílých Karpat. Preslia* 47: 129-144.

Valachovič, M. 1999. *Centrálna databáza fytocenológických zápisov (CDF) na Slovensku.* In: Zborník 7. zjazdu SBS, pp. 218-220. Hrabušice, Podlesok, SK.

Varga, Z., Varga-Sipos, J., Orci, M.K. & Rácz, I. 2000. Felszáraz gyepek az Aggteleki karszton. In: Virágh, K. & Kun, A. (eds.) *Vegetáció és Dinamizmus,* pp. 195-238. MTA ÖBKI, Vácrátót, HU.

Wagner, H. 1941. Die Trockenrasengesellschaften am Alpenostrand. Eine Pflanzensoziologische Studie. *Denkschr. Akad. Wiss. Wien., Math.-Nat. Kl.* 104: 1-81.

Willems, J.H. 1982. Phytosociological and geographical survey of *Mesobromion* communities in Western Europe. *Vegetatio* 48: 227-240.

Zar, J.H. 1999. *Biostatistical analysis*. 4th. ed. Prentice & Hall, Upper Saddle River, NY, US.

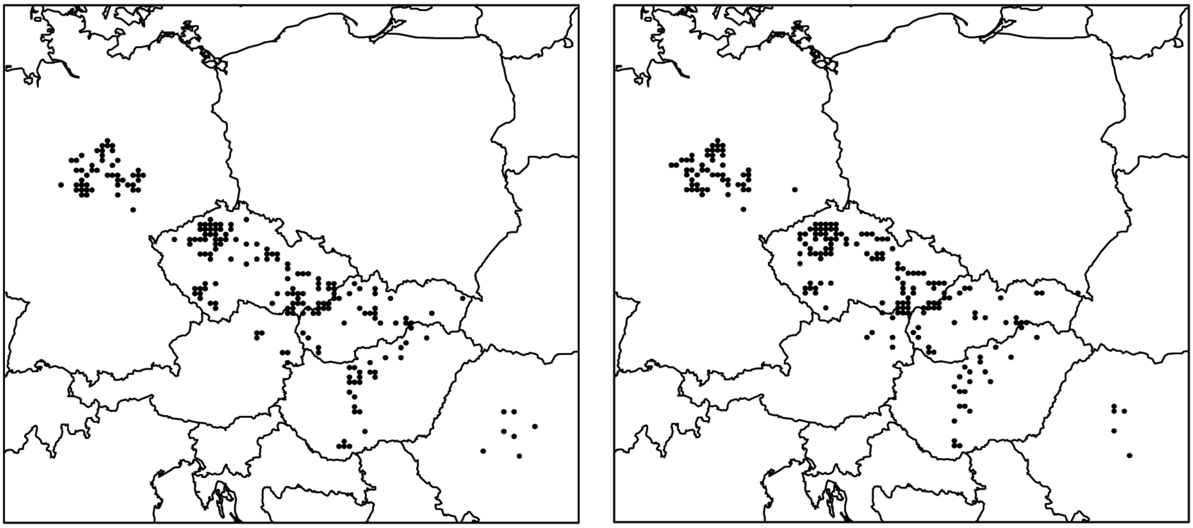
Zólyomi, B. & Fekete, G. 1994. The Pannonian loess steppe: differentiation in space and time. *Abstr. Bot.* 18: 29-41.

Received 13 July 2006;
Accepted 15 March 2007;
Co-ordinating Editor: P. Dixon.

For App. 1-2, see also JVS/AVS Electronic Archives; www.opuluspress.se/

App. 1. Geographic distribution of the relevés in the TRAINING and TEST data sets. Each point represents 1-12 relevés.

A. TRAINING data set B. TEST data set

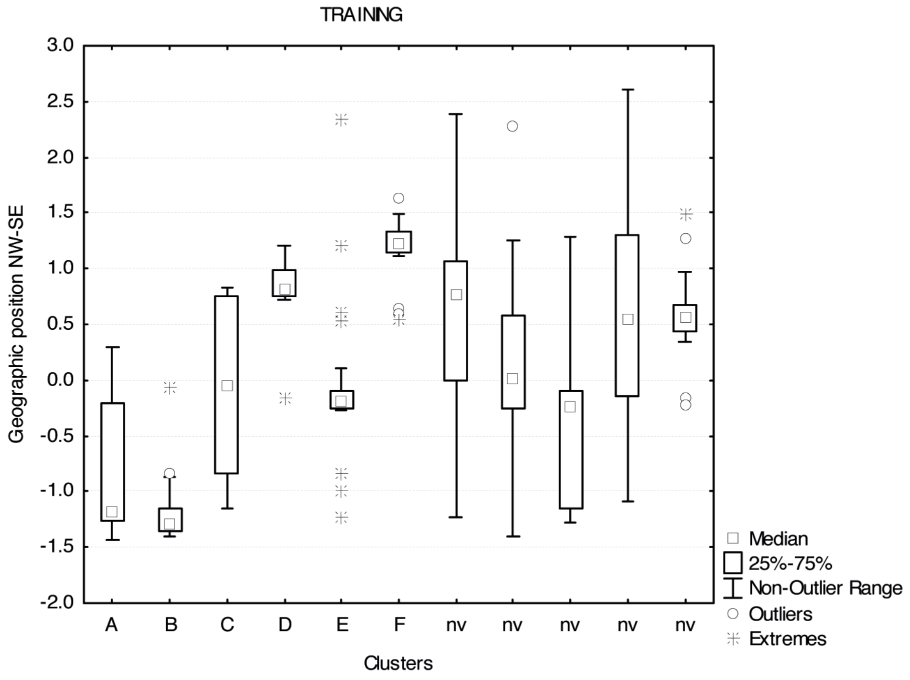


App. 1-2. Internet supplement to: Illyés, E.; Chytrý, M.; Botta-Dukát, Z.; Jandt, U.; Škodová, I.; Janišová, M.; Willner, W. & Hájek, O. 2007. Semi-dry grasslands along a climatic gradient across Central Europe: Vegetation classification with validation. *J. Veg. Sci.* 18: 835-846.

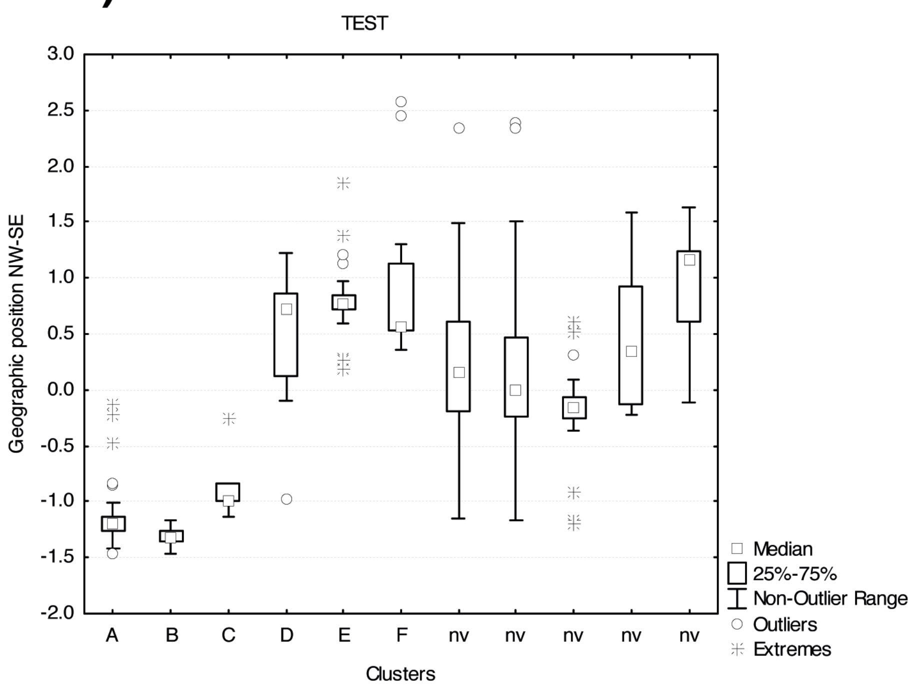
OPULUSPRESS

App. 2. Geographic distribution along the NW-SE gradient of the clusters of the partitions of (**A**) TRAINING and **(B)** TEST data sets at the level of 11 clusters. Lower position on vertical axis represents a more NW distribution, higher position a more SE distribution. Letters A-F label corresponding valid clusters in TRAINING and TEST data sets; nv indicates non-valid clusters.

A)



B)



OPULUSPRESS