SPECIAL FEATURE: VEGETATION SURVEY

Surveying Southern Mongolia: application of multivariate classification methods in drylands with low diversity and long floristic gradients

Karsten Wesche & Henrik von Wehrden

Wesche, K. (corresponding author, karsten.wesche@senckenberg.de): Botany Department, Senckenberg Museum of Natural History Görlitz, P.O. Box 300154, 02806 Görlitz, Germany

Von Wehrden, H. (vonwehrd@leuphana.de): Institute of Ecology (Faculty of Sustainability) & Centre of Methods, Leuphana University, 21335 Lüneburg, Germany.

Abstract

Question: How do standard multivariate classification algorithms perform for Mongolian dryland vegetation characterized by low diversity and very long floristic gradients?

Location: Southern Mongolian Gobi.

Methods: We compared several widely used classification methods based on 1231 relevés obtained with a similar sampling method. We applied agglomerative cluster algorithms (flexible beta/FLEX, average linkage/UPGMA, weighted linkage/WPGMA, complete linkage, Ward's clustering) and a divisive classification technique (TWINSPAN); data were reduced to presence/absence. We compared results against a published phytosociological classification (PHYTO), against environmental background data, and with respect to the presence of significant indicator species.

Results: Complete linkage was inferior to other methods. TWINSPAN, UPGMA, flexible beta and WPGMA gave partly similar clusters, with FLEX and WARD showing the highest pair-wise similarity. Classifications of all methods except CL partly agreed with PHYTO classification. Clusters of all methods had significant indicator species, but Ward's method had the highest number of indicator species, followed by the PHYTO classification and FLEX, TWINSPAN, UPGMA and WPGMA. The latter four methods all yielded clusters that differed in terms of precipitation, but TWINSPAN, FLEX and Ward's method performed best under this criterion. PHYTO and CL ranked last in partitioning the precipitation gradient. Comparisons with ordinations indicated that classification algorithms capture the main floristic gradient but were less successful than the phytosociological approach to elucidate the finer structures.

Conclusion: Performance of classification methods differed depending on the applied validation approach and we thus caution against uncritically adopting a single evaluation/validation criterion. Most numerical approaches can aid sorting of large data sets, while details of manual syntaxonomic classifications are not easily reproduced. Choice of the most appropriate classification and validation method thus clearly depends on the overall aim of a given study.

Introduction

Vegetation classification has been a key activity in vegetation science for more than a century (Kent & Ballard 1988; Dierschke 1994). This is related to its high relevance for applied questions such as vegetation mapping; a topic that has received considerable interest in light of global climate and land-use change (Bohn et al. 2000; Xie et al. 2008). Classification is also highly relevant for nature conservation (Lengyel et al. 2008).

Classification has been the subject of one of the most intensive methodological discussions in vegetation science. The traditional phytosociological approach

(Braun-Blanquet 1964; Mueller-Dombois & Ellenberg 1974) has received severe criticism in the last decades, which witnessed the advent of numerical classification (Kent & Coker 1992). The main objection was the semi-formal, non-numerical classification process itself, which relied heavily on non-statistical table work. Another objection was the circularity in designating character species for classes of vegetation samples; this is now, however, seen as an iterative problem that can be handled mathematically (Bruelheide 2000). In spite of recent attempts to re-formulate phytosociology in the framework of more strictly formalized rules (Dengler 2003), most vegetation scientists would probably prefer numerical classification techniques.

Enthusiasm for numerical classification was high in the 1970s and 1980s, when a large number of algorithms and specific modifications were introduced (Kent & Ballard 1988; Gauch 1994). Developments were less diverse in the 1990s, with some contributions from Central Europe being notable exceptions (Bruelheide 2000; Chytrý et al. 2002; Tichý 2002). More recently, methodological discussions seem to have been revived (e.g. Aho et al. 2008; van Tongeren et al. 2008; Roleček et al. 2009; Tichý et al. 2010), and still no consensus on the most suitable methods has been reached. The political and legal framework with its focus on phytosociologically derived habitat classification, as well as research traditions, may explain why most of the mentioned recent methodological contributions come from Central and Western Europe; which is also where most methods have been tested with real data sets so far (Bruelheide 2000; Wesche et al. 2005b).

Here, we present a multivariate classification of vegetation types in the dry regions of the Gobi in southern Mongolia. These drylands are characterized by a low productivity, have a limited overall species pool, often low alpha diversity and long environmental and associated floristic gradients, ranging from almost bare deserts to forests within less than 300-km distance (von Wehrden & Wesche 2007). Relevés had to be classified for vegetation mapping (Kaczensky et al. 2008), and as the benchmark texts on Mongolian vegetation were based on phytosociological approaches (Hilbig 1995, 2000), we also initially applied phytosociological table work (von Wehrden et al. 2009b). Here, we re-classified the data using methods that are frequently employed in vegetation classification (McCune et al. 2002; Aho et al. 2008).

We compared results from these classifications against each other and against the available phytosociological scheme. Results of all classifications were also plotted on ordinations, a widely recommended form of complementary analysis (Kent & Ballard 1988; Legendre & Legendre 1998). We also tested the derived classes for statistically significant differentiation, and also for the presence of indicator species that are statistically associated with a given class (Dufrêne & Legendre 1997; Tichý et al. 2010). This approach is insofar analogous to phytosociology as it emphasises the presence of a limited number of characteristic species that have a high fidelity to the given class. Finally, we tested the derived classes for differences with respect to abiotic background information. The idea that communities should represent distinct habitat conditions dates back to the early phases of vegetation science (Flahault & Schröter 1910). We focused on precipitation because that is the main controlling factor in this arid to semi-arid vegetation (Noy-Meir 1973), which has been confirmed by univariate and multivariate gradient analysis of the present data (von Wehrden & Wesche 2007; von Wehrden et al. 2009c). Specifically, we addressed the following questions:
(1) Is there any correspondence between multivariate classification methods, and between statistical methods and the available phytosociological scheme?
(2) Are derived classes characterized by distinct indicator species?
(3) Can classes be related to distinct abiotic conditions, and does the general structure of the classification reflect the known large-scale gradients in the data?

Our study does not aim at a fundamental test of available methods, which are usually performed on artificial data sets with known properties. Instead, we add another example to the limited number of real-world comparative studies.

Study Sites

Appendix S1 gives an overview of the study sites in southern Mongolia. All sites are characterized by warm summers, cold winters and generally low precipitation. There are strong climatic gradients, with mountains naturally being colder and receiving more precipitation (von Wehrden & Wesche 2007). In addition, our sampling covered principally different precipitation regimes and biogeographical regions: the westernmost regions mediates to Mid-Asian conditions, where winter precipitation occurs frequently, while the Trans-Altay Gobi (von Wehrden et al. 2006a) and – to a lesser extent – the Gobi Altay (Wesche et al. 2005a) are typical for the arid, highly continental regions of Central Asia that receive hardly any snow. The Eastern Gobi, in contrast, mediates to the much more summer-humid conditions of monsoonal Northeast Asia.

The precipitation gradient from 40 to $> 200\,mm$ (Appendix S1) represents the transition from true deserts with, at most, contracted and often annual vegetation occurring only in exceptionally wet years, over more or

less diffuse steppe vegetation, to completely closed mats and even forests (von Wehrden et al. 2009b). Our data set comprised 15 associations/communities (see syntaxonomical overview in Appendix S2). Desert steppes and other semi-desert vegetation of the *Stipetea glareoso-gobicae* and grass steppes of the *Cleistogenetea squarrosae* are the most common community groups. Saline sites and related habitats were excluded as they hardly shared any species with the remaining stands and were thus, not surprisingly, placed widely apart in all manual and numerical approaches.

Methods

Fieldwork and data processing

Relevés were 10×10 m in size, which was large enough to capture the floristic composition. Positions were geo-referenced with hand-held GPS, which allowed extraction of data on topography, elevation, remotely sensed productivity, extrapolated precipitation and temperature from standard models (details in von Wehrden & Wesche 2007). Site selection was deliberate but aided by unsupervised classifications of Landsat ETM+ data (von Wehrden et al. 2006b). Samples were selected to represent all major vegetation types, including extrazonal and azonal sites.

Originally collected cover values were reduced to presence–absence. Cover values vary hugely from year to year under the highly variable precipitation regime of southern Mongolia. Pronounced changes in species cover values may thus be caused by inter-annual variability rather than by lasting differences in plant community composition. We used presence/absence data since the previously published phytosociological vegetation classification (PHYTO) was also solely based on presence/absence data.

Statistical analysis

The data were subjected to six numerical classification methods commonly applied in vegetation science (Aho et al. 2008), where properties are relatively well known (Podani 2000; McCune et al. 2002). We neglected more recent proposals and also did not use non-hierarchical approaches (Bruelheide 2000; Černá & Chytrý 2005; De Cáceres et al. 2010). Testing these methods would, however, be an interesting goal of another study. We used three agglomerative cluster techniques that tend to yield larger equal-sized clusters (complete linkage/CL – space dilatating; flexible beta/FLEX; Ward's method/WARD – space conserving), and two methods that are intermediate (average linkage/UPGMA – space conserving; weighted average linkage/WPGMA – space conserving). Single linkage was not considered because of its well-known shortcomings (Legendre & Legendre 1998; Podani 2000;

McCune et al. 2002). In order to keep methods comparable, we used Sørensen dissimilarities in all cluster applications, but applied a square-root transformation of the dissimilarity matrix as this is known to improve metric properties (Gower & Legendre 1986). Sørensen is identical to the Bray-Curtis dissimilarity if applied to presence/absence data, and represents the most widely recommended dissimilarity measure in (vegetation) ecology (McCune et al. 2002). The second most important measure, Euclidean distance, was not applicable because it results in serious distortions if applied to data sets with a high number of zeros (Legendre & Legendre 1998). There are numerous other dissimilarity indices, but testing dissimilarities was again not our main focus.

Using Sørensen poses a problem with Ward's clustering, which considers variances and should thus be based on a metric distance (Euclidean). WARD is akin to the non-hierarchic k-means clustering, which also optimizes a variance criterion. Recently, more flexible extensions of k-means clustering were published, but these have so far not been widely applied (Podani 2005). Instead, we used flexible beta clustering with a value of $\beta = -0.25$, which should perform largely similar to WARD but is not limited to Euclidean distance (McCune et al. 2002). Cluster dendrograms were pruned to yield between one and 60 groups.

TWINSPAN is still the most widely used divisive numerical classification technique in vegetation science. It is based on the chi-square dissimilarities and is also capable of handling sparse data. Its shortcomings have been discussed in a huge body of literature, yet many vegetation scientists seem to adopt a pragmatic approach and cherish its often-useful results. Here, we used a recent modification of the original algorithm that accounts for decreasing β-diversity with progressive cluster division (Roleček et al. 2009).

Our first comparison was to assess the relation between our phytosociological classification and the cluster results. We tested each cluster method at each group size against our manually derived 15 classes using the Rand index in a chance-corrected version that accounts for increasing similarities with increasing numbers of clusters (Vinh et al. 2009). Although often used in statistical science (Legendre & Legendre 1998), applications are rather sparse in vegetation ecology (e.g. De Cáceres et al. 2010). We also compared all classification methods against each other, using the Rand index and calculating it for the level of 15 groups each.

Most of our methods are more or less space conserving, so just calculating the cophenetic correlations would not be very informative. Classification quality can also be assessed based on the number of species that are restricted to a given cluster. A recent development in the field is

OptimClass (Tichý et al. 2010), which is based on measures of fidelity. The number of applications for this new method is still limited and in our case the relatively low computational demand, one of the main advantages of OptimClass as discussed by Tichý et al. (2010), was not relevant because all employed methods ran relatively rapidly on our standard computer. We opted for standard indicator species analysis (ISA) that can also be used for this type of problem. ISA is very widely used and also has modest computational requirements (Dufrêne & Legendre 1997). We additionally assessed within-cluster dissimilarity versus overall dissimilarity using multiple response permutation procedure (MRPP, see McCune et al. 2002), which can be used to test differences between groupings. MRPP provides a measure of differentiation among clusters.

Finally, we tested clusters against the mean annual precipitation derived from the WorldClim data set (Hijmans et al. 2005), which has proven useful in our working area (von Wehrden & Wesche 2007). One-way ANOVAs were calculated for each classification method at each level of pruning, and F-values from each ANOVA were then plotted against the number of groups. Rand indices and numbers of indicator species were displayed in a similar manner. DCA ordinations were used to estimate species turnover rates (von Wehrden et al. 2009a) and for the visualization of the position of clusters in floristic space (PCoAs based on Sörensen dissimilarity are given in the Supporting information). Ellipses of the classifications were plotted on the ordination space based on the standard deviation of the average of the sample scores in a given class. With the exception of TWINSPAN, all analyses were calculated with customized loops in the R framework (R Development Core Team 2010, Vienna, Austria) using packages vegan, cluster, MASS, labdsv, asbio, stats, car and e1071.

Results

The overall number of species was 421 and α-diversity was low (plot-level mean 9.4, SD = 6.0), which yielded a high Whittaker's beta at 45.8. DCA also indicated a huge β-diversity: the first axis had a gradient length of 11.3 SD units (Appendix S3). Axis 1 correlated closely with altitude, precipitation, productivity and diversity (Appendices S3 and S4). The second axis differentiated between different types of semi-desert and desert vegetation on arid sites. Most phytosociological units were reasonably distinct in the ordination plot, although there were partly pronounced overlaps in the two-dimensional space. Classes derived from numerical classification were more clumped in the DCA (Appendix S5). CL differentiated classes mainly along the primary gradient of the first axis, while FLEX, UPGMA, WARD, WPGMA and TWINSPAN also differentiated some groups along the second axis.

According to the corrected Rand index, results of numerical classifications were partly similar to the PHYTO classification (Fig. 1). Similarity in the pair-wise comparison reached a plateau at ca. 15–25 clusters corresponding to a Rand index of ca. 0.25–0.30 for UPGMA and WPGMA, while WARD and FLEX peaked at eight to ten clusters and declined thereafter. TWINSPAN peaked at a low number of five clusters (Rand index almost 0.3), yet showed a slight decline for more than six groups, levelling at around 0.27. Values for CL remained distinctly lower throughout, and CL also proved different in the pair-wise comparison (Table 1), showing low similarity values to all other methods. As expected, FLEX and WARD had the largest similarity, with UPGMA and WPGMA showing moderate similarities to all methods except CL. TWINSPAN most closely resembled FLEX and WARD.

Most cluster algorithms yielded mainly larger clusters, with relatively few clusters comprising single-relevé
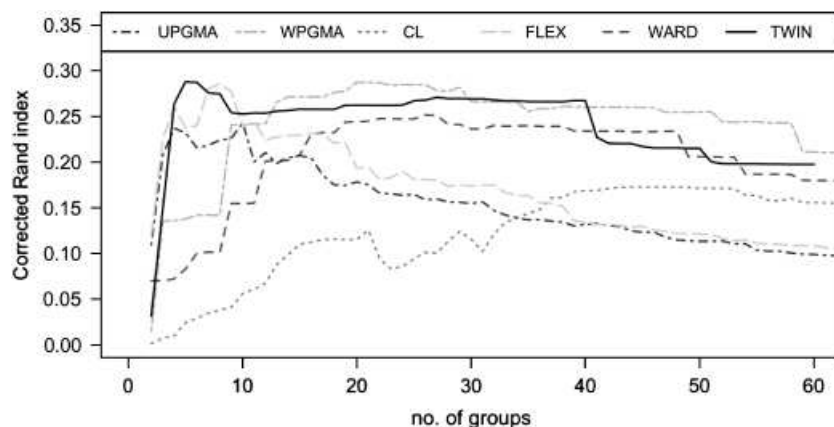

Fig. 1. Corrected Rand index comparing the classification results of the numeric classification against the phytosociological classification. Note that this index corrects for the number of groups and therefore gives relatively low values for fine partitions.

groups (Appendix S7). Chains of singletons were totally absent in WARD, FLEX and CL, as they were in PHYTO. Performance of UPGMA, WPGMA and TWINSPAN was intermediate, with a limited tendency to produce singletons at finer partitions.

The numbers of significant indicator species showed different trends (Fig. 2). For partitions with > 10 clusters, CL, UPGMA and WPGMA showed about 130–200 indicator species. TWINSPAN was initially comparable but numbers of indicators dropped with increasing partitions. WARD outperformed all other methods under this criterion, yielding 200–280 species. FLEX, which was otherwise similar to WARD in other criteria, was less powerful here and resembled results from CL. PHYTO classification with 15 classes yielded 182 significant indicators, and thus performed slightly better than TWINSPAN, UPGMA and WPGMA and somewhat worse than CL and FLEX.

We also correlated the number of significant indicator species with their summed indicator values for all classes at each clustering level (Fig. 3). FLEX and WARD showed consistently high positive correlations. This was followed by PHYTO, while correlations for all other methods were low beyond the first ten to 15 groups. Thus groups with more significant indicators do not necessarily have higher sums of indicator values, suggesting that many of the respective indicator species have limited discriminatory power.

In terms of the ANOVA-based environmental differentiation, TWINSPAN, WARD, UPGMA and WPGMA did not show major differences in F-values at more than 20 groups (Fig. 4). CL showed rather weak partitioning of the environmental information based on the cluster groups. Notably, WARD started with the initially highest F-values yet levelled to a similar performance as UPGMA and WPGMA at more than ten groups. The F-value for the PHYTO classification at 15 units was 95, showing that phytosociologically derived clusters were similarly distinct in terms of precipitation as those from CL, and less distinct than clusters derived with the remaining techniques. A more detailed visualization of the environmental space of all classifications using boxplots (Appendix S8) for the level corresponding to the syntaxonomic classification ($n = 15$) reveals that CL performs poorly, while all other methods show a pronounced differentiation with respect to the main environmental factor: precipitation. MRPP also indicated that WARD, FLEX and TWINSPAN were good in finding reasonably well-differentiated clusters, especially at coarser levels (Appendix S9). In this respect, UPGMA and CL performed worse than other methods, and groups from PHYTO were also moderately distinct according to MRPP.

Discussion

Comparison of five agglomerative, one divisive algorithm and one manual classification method (phytosociology) suggest that standard numerical classification yields meaningful results even in data sets with extremely long floristic gradients combined with relatively low diversity. Algorithms gave acceptable results with respect to all main aspects assessed here; namely similarity to each other and an available phytosociological classification (exception CL), number of significant indicator species
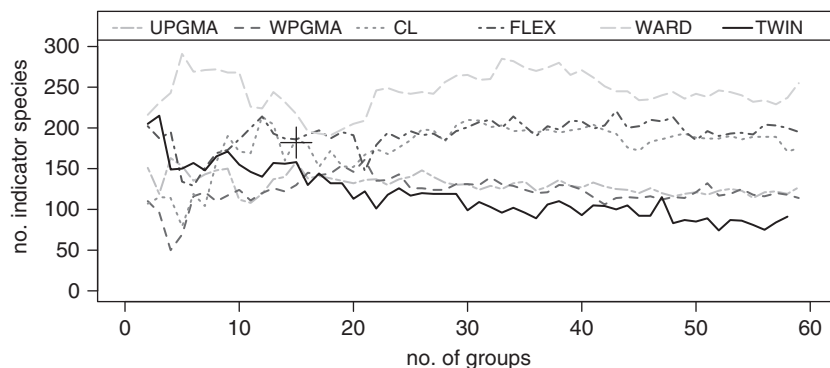
Table 1. Pair-wise comparison (corrected Rand index) of the various classification methods. For simplicity, values were calculated for $n = 15$ only, which corresponds to the number of groups in the phytosociological classification, where no finer classification was available.

	WPGMA	CL	FLEX	WARD	TWIN	PHYTO
UPGMA	0.24	0.33	0.21	0.19	0.21	0.20
WPGMA	–	0.07	0.31	0.31	0.26	0.27
CL	–	–	0.05	0.04	0.11	0.09
FLEX	–	–	–	0.62	0.41	0.23
WARD	–	–	–	–	0.38	0.20
TWIN	–	–	–	–	–	0.25

Fig. 2. Number of significant species according to an indicator species analysis based on the given numerical classification (999 permutations). The large cross at 15 groups indicates the number for the phytosociological classification.
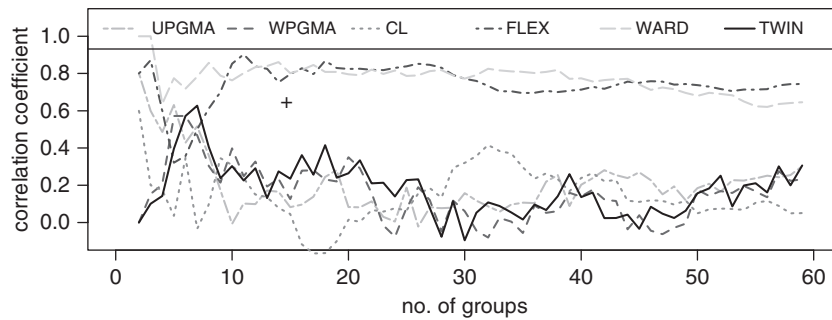
Fig. 3. Spearman rank correlation between the number of significant indicator species per cluster group and their summed indicator value per group. The large cross at 15 groups gives the value for the phytosociological classification.
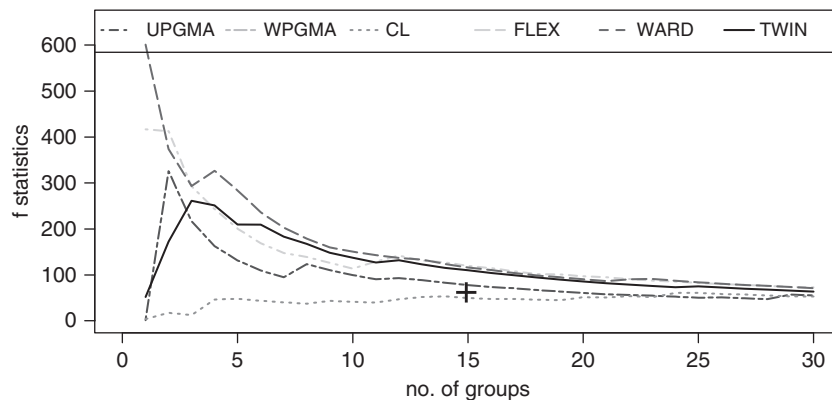

Fig. 4. *F*-statistic as derived from ANOVAs for each group level for the given numerical classification. Note that only groups two to 30 are shown. The large cross at 15 groups gives the value for the phytosociological classification.

and distinctiveness in environmental space. Formation of singletons occurred but was not pronounced even in UPGMA, WPGMA and TWINSPAN, as intermediate methods. Most methods performed similarly for the finer partitions, which is not surprising since they are mostly agglomerative and thus similar in the first fusion steps. We thus concentrated on the larger groupings that are relevant for vegetation classification.

Application of numerical classifications is complicated by the multitude of choices available. Our results imply that presence/absence data seem to work well, which is not surprising. Beal's smoothing is an option for sparse presence/absence data (McCune 1994; De Cáceres & Legendre 2008), but given the low amount of covariance in our data it did not seem appropriate. Other measures are available, but we wanted to restrict comparisons to the modification of the classification algorithm, and also advocate using standard distance measures unless there are good reasons for other choices. Transformations, as applied here on the dissimilarity matrix, may be useful, but our results confirm that they usually have a stronger impact on branch lengths than on cluster topology (McCune et al. 2002).

Comparison with the phytosociological classification

Vegetation science in Mongolia benefits from the availability of the benchmark descriptions of Hilbig (1995, 2000). When we classified our own data set with phytosociological procedures, most relevés fell in units already described by Hilbig. The corrected Rand index suggests that statistically derived clusters support several of the phytosociological units. This is not necessarily expected, given that the cluster algorithms are based on a dissimilarity matrix and thus the total species set, while phytosociology is essentially based on the presence of characteristic species with a high fidelity (Kent & Coker 1992). We initially tried to validate our PHYTO classification with more formal assessments of fidelity (COCKTAIL, see Bruelheide 2000; Chytrý et al. 2002), which failed for all but the alpine/montane stands, which have a high number of associated species (data not shown, see syntaxonomic tables in von Wehrden et al. 2009b), while the other vegetation types did not have enough abundant and statistically associated species.

The relatively lax criteria for fidelity applied in standard phytosociology (Dengler 2003) still allow classification of

samples under such conditions, and approaches based on the full species assemblage also work well in these sparse data matrices. Close correspondence between cluster analyses/TWINSPAN and phytosociological units has been repeatedly observed in the past (e.g. van der Maarel 1979; Wesche et al. 2005b; Dúbravková et al. 2010).

In our particular case, UPGMA and the closely related WPGMA seem to resemble the original classification better than CL (at least initially), FLEX or WARD. These were also particularly good in differentiating classes along the main floristic gradient captured by ordination methods, which in this case corresponds to the transition from arid desert vegetation to dry steppe and to mountain steppe and alpine swards. Indeed, UPGMA has been recommended as the most widely applicable algorithm, at least if the numbers of classes are high (McCune et al. 2002; Tichý et al. 2010). TWINSPAN performed almost as well, and also had especially high correspondence to our syntaxonomic classification at the initial partitions. Units within the arid (semi-) desert vegetation, which are differentiated mainly on the second axis of DCA, were less clearly differentiated by all numerical classification methods (Appendix S5). It is well known that detrending may introduce artificial distortions along the second axis, but in our case the *Iljinia regelii* community and the *Calligono-Haloxyletum* (lower part of axis 2) were indeed floristically distinct from the *Nitrario-Kalidetum* and the *Salsolo-Reaumurietum* (upper part), which tend to have higher salinity. The character species approach of phytosociology seems superior in this respect.

Presence of indicator species

All classifications including PHYTO were supported by a number of significantly associated indicator species (up to 30 clusters) and thus can all be considered valid. There was no clear unimodal behaviour with increasing number of clusters, which has been demonstrated for other data sets (Tichý et al. 2010), suggesting that finer partitions are still interesting. At finer partitions, however, indicator species seem to have lower indicator values, as shown in Fig. 2, suggesting that the discriminatory power of these indicators is weak.

The limited performance of TWINSPAN with respect to ISA may seem surprising. Its underlying indicator species concept is, however, based on gradient analysis, with the result that indicators are only valid for the given floristic context, i.e. the respective classification step (Legendre & Legendre 1998). This is influenced by the behaviour of species in the underlying correspondence analysis, which is known to misplace very rare species (van Groenewoud 1992).

Unlike the Rand index, the number of significant indicator species pointed to WARD as the most suitable algorithm. This, FLEX and also CL tend to yield large clusters, which are apparently well supported under this criterion. Moreover, WARD gives a high number of significant indicators even at the upper levels of the cluster hierarchy, suggesting that coarser partitions are better supported than in other methods – a desirable feature for vegetation scientists who are interested in groupings on upper hierarchical levels. MRPP also supported the view that WARD and FLEX yield relatively well-differentiated clusters, while those from PHYTO classification were only moderately distinct.

Differences in environmental space

Comparisons against climatic background data are much more common in gradient analysis than in classification (but see Carleton et al. 1996), although groupings can also be compared using discriminant analysis, (Per)MANOVA or related techniques. In all multivariate analyses, assessments depend on structurally informative environmental data, which are often not available in large-scale vegetation surveys. In our case, DCA revealed precipitation as one of the most important predictors, which is typical for drylands, rendering the significant differentiation of cluster with respect to precipitation not surprising.

Magnitude of *F*-values varied among approaches. Most cluster algorithms performed more or less similarly, at least if finer partitions were considered; CL performed poorly at finer levels and resembled other algorithms at coarser levels. FLEX and WARD outperformed all other methods in the coarser groupings, yet approximated the average performance at partitions of more than nine groups. The PHYTO classification performed worse than the best numerical classifications. Phytosociologists pay high attention to characteristic species, which may represent unique azonal stands and site conditions even under a similar overall climate and general species composition. An example in our data set is beds of intermittent streams that host a limited number of specialized and often not very common disturbance indicators, but at the same time comprise a number of species from the surrounding vegetation. In phytosociology, these samples were classified as distinct, while classification algorithms at least partly emphasized the similarity to surrounding vegetation. Both ordination (Appendix S3) and boxplots (Appendix S8) support this view. This reasoning would call for a more complex analysis in multivariate environmental space. Environmental data are, however, of different quality and have a different spatial resolution, so we refrained from using multivariate tests such as MANOVA.

Conclusions

Our analyses suggest that standard multivariate classification procedures seem to work even under the special conditions of species-poor and highly variable drylands. The applied validations and comparisons indicated generally reasonable performance of all methods, but results of different validation methods differed in detail. Performance of algorithms varied under different criteria, but flexible beta and UPGMA work well with respect to most of the applied criteria. Flexible beta offers an alternative to the often-used WARD algorithm, but unlike WARD is not restricted to the not always applicable Euclidean distances. Flexible beta should thus be used more often in vegetation classification. Although assumptions were violated, WARD nonetheless performed well under a number of criteria. We see no apparent reason to use specialized approaches (both in terms of dissimilarity measure and clustering strategy) even in the extreme case of the Gobi. Combination of different approaches does, however, reveal the strengths and weaknesses of different algorithms, since the coarser divisions may be more reliably captured by some methods, while finer divisions may be superiorly performed by other techniques. In our case, phytosociology was partly similar to numerical methods and performed best for azonal vegetation.

Taken together, we thus caution against uncritically adopting a single evaluation/validation criterion for assessing classification performance. Different measures yield different results, and suitability of both classification and validation depends on the given scientific aims. Close inspections of the raw data, be it with table work or ordinations, are in any case instructive. Further research into the combination of different methods thus remains desirable, also beyond the drylands.

Acknowledgements

References

Aho, K., Roberts, D. & Weaver, T. 2008. Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. *Journal of Vegetation Science* 19: 549–562.

Bohn, U., Gollub, G. & Hettwer, C. eds. 2000. *Map of the natural vegetation of Europe*. Bundesamt für Naturschutz, Bad Godesberg, Bonn, DE.

Braun-Blanquet, J. 1964. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Springer, Berlin, DE.

Bruelheide, H. 2000. A new measure of fidelity and its application to defining species groups. *Journal of Vegetation Science* 11: 167–178.

Carleton, T.J., Still, R.H. & Nieppola, J. 1996. Constrained indicator species analysis (COINSPAN): an extension of TWINSPAN. *Journal of Vegetation Science* 7: 125–130.

Ĉerná, L. & Chytrý, M. 2005. Supervised classification of plant communities with artificial neural networks. *Journal of Vegetation Science* 16: 407–414.

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

De Cáceres, M. & Legendre, P. 2008. Beals smoothing revisited. *Oecologia* 156: 657–669.

De Cáceres, M., Font, X. & Oliva, F. 2010. The management of vegetation classification with fuzzy clustering. *Journal of Vegetation Science* 21: 1138–1151.

Dengler, J. 2003. *Entwicklung und Bewertung neuer Ansätze in der Pflanzensoziologie unter besonderer Berücksichtigung der Vegetationsklassifikation*. Martina Galunder, Nümbrecht, DE.

Dierschke, H. 1994. *Pflanzensoziologie: Grundlagen und Methoden*. 1st ed. Ulmer, Stuttgart, DE.

Dúbravková, D., Chytrý, M., Willner, W., Illyés, E., Janišová, M. & Kállayné Szerényi, J. 2010. Dry grasslands in the Western Carpathians and the northern Pannonian Basin: a numerical classification. *Preslia* 82: 165–221.

Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.

Flahault, C. & Schröter, C. (eds.). 1910. *Phytogeographische Nomenklatur. Berichte und Vorschläge*. 3. Congrès Int. Bot. Bruxelles 14–22. Zürcher & Furrer, Mai 1910, Zürich, CH.

Gauch, H.G. 1994. *Multivariate analysis in community ecology*. 7th ed. Cambridge University Press, Cambridge, UK.

Gower, J.C. & Legendre, P. 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* 3: 5–48.

Gubanov, I.A. 1996. *Conspectus of the Flora of Outer Mongolia (Vascular Plants) (in Russian)*. Valang Publishers, Moscow, RU.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Hilbig, W. 1995. *The vegetation of Mongolia*. SPB Academic Publishing, Amsterdam, NL.

Hilbig, W. 2000. Kommentierte Übersicht über die Pflanzengesellschaften und ihre höheren Syntaxa in der Mongolei. *Feddes Repertorium* 111: 75–120.

Kaczensky, P., Ganbaatar, O., von Wehrden, H. & Walzer, C. 2008. Resource selection by sympatric wild equids in the Mongolian Gobi. *Journal of Applied Ecology* 45: 1762–1769.

Kent, M. & Ballard, J. 1988. Trends and problems in the application of classification and ordination methods in plant ecology. *Vegetatio* 78: 109–124.

Kent, M. & Coker, P. 1992. *Vegetation description and analysis – A practical approach.* 1st ed. Belhaven Press, London, UK.

Legendre, P. & Legendre, L. 1998. *Numerical Ecology.* Elsevier, Amsterdam, NL.

Lengyel, S., Déri, E., Varga, Z., Horváth, R., Tóthmérész, B., Henry, P.-Y., Kobler, A., Kutnar, L., Babij, V., Seliškar, A., Christia, C., Papastergiadou, E., Gruber, B. & Henle, K. 2008. Habitat monitoring in Europe: a description of current practices. *Biodiversity and Conservation* 17: 3327–3339.

McCune, B. 1994. Improving community analysis with the Beals smoothing function. *Ecoscience* 1: 82–86.

McCune, B., Grace, J.B. & Urban, D.L. 2002. *Analysis of ecological communities.* MjM Software Design, Gleneden Beach, US.

Mueller-Dombois, D. & Ellenberg, H. 1974. *Aims and methods of vegetation ecology.* 1st ed. John Wiley & Sons, New York, NY, US.

Noy-Meir, I. 1973. Desert ecosystems: environment and producers. *Annual Review of Ecology and Systematics* 4: 25–41.

Podani, J. 2000. *Introduction to the exploration of multivariate biological data.* Backhyus , Leiden, NL.

Podani, J. 2005. Multivariate exploratory analysis of ordinal data in ecology: pitfalls, problems and solutions. *Journal of Vegetation Science* 16: 497–510.

Roleček, J., Tichý, L., Zelený, D. & Chytrý, M. 2009. Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity. *Journal of Vegetation Science* 20: 596–602.

Tichý, L. 2002. JUICE, software for vegetation classification. *Journal of Vegetation Science* 13: 451–453.

Tichý, L., Chytrý, M., Hájek, M., Talbot, S.S. & Botta-Dukát, Z. 2010. OptimClass: using species-to-cluster fidelity to determine the optimal partition in classification of ecological communities. *Journal of Vegetation Science* 21: 287–299.

van der Maarel, E. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* 39: 97–114.

van Groenewoud, H. 1992. The robustness of correspondence, detrended correspondence, and TWINSPAN analysis. *Journal of Vegetation Science* 3: 239–246.

van Tongeren, O., Gremmen, N. & Hennekens, S. 2008. Assignment of relevés to pre-defined classes by supervised clustering of plant communities using a new composite index. *Journal of Vegetation Science* 19: 525–536.

Vinh, N.X., Epps, J. & Bailey, J. 2009. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? Proceedings of the 26th International Conference on Machine Learning, Montreal, CA.

von Wehrden, H. & Wesche, K. 2007. Relationships between climate, productivity and vegetation in southern Mongolian drylands. *Basic and Applied Dryland Research* 1: 100–120.

von Wehrden, H., Hilbig, W. & Wesche, K. 2006a. Plant communities of the Great Gobi A special protected area in the Mongolian Transaltay Gobi. *Feddes Repertorium* 117: 526–570.

von Wehrden, H., Wesche, K., Reudenbach, C. & Miehe, G. 2006b. Vegetation mapping in Central Asian dry ecosystems using Landsat ETM+. A case study on the Gobi Gurvan Saykhan National Park. *Erdkunde* 60: 261–272.

von Wehrden, H., Hanspach, J., Bruelheide, H. & Wesche, K. 2009a. Pluralism and diversity: trends in the use and application of ordination methods 1990–2007. *Journal of Vegetation Science* 20: 695–705.

von Wehrden, H., Wesche, K. & Miehe, G. 2009b. Plant communities of the southern Mongolian Gobi. *Phytocoenologia* 39: 331–376.

von Wehrden, H., Zimmermann, H., Hanspach, J., Ronnenberg, K. & Wesche, K. 2009c. Predicting plant species and communities by using GIS and Landsat data in a southern Mongolian mountain range. *Folia Geobotanica* 44: 211–225.

Wesche, K., Miehe, S. & Miehe, G. 2005a. Plant communities of the Gobi Gurvan Sayhan National Park (South Gobi Aimag, Mongolia). *Candollea* 60: 149–205.

Wesche, K., Partzsch, M., Krebes, S. & Hensen, I. 2005b. Gradients in dry grassland and heath vegetation on rock outcrops in eastern Germany – an analysis of a large phytosociological data set. *Folia Geobotanica* 40: 341–356.

Xie, Y., Sha, Z. & Yu, M. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology* 1: 9–23.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Map of study area and sampling sites.

Appendix S2. Overview of phytosociological syntaxonomy.

Appendix S3. DCA of the entire data set, overlaid by the phytosociological classification.

Appendix S4. PCoA of the entire data set, overlaid by the phytosociological classification.

Appendix S5. Overlays of numerical classifications on DCA ordination.

Appendix S6. Overlays of numerical classifications on PCoA ordination.

Appendix S7. Relationship between number of groups and groups not comprised of singletons.

Appendix S8. Boxplots of precipitation plotted against numerically derived classifications and the phytosociological classification.

Appendix S9. Results of multiple response permutation procedure.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.