UDC 581.55:311(497)
Original scientific paper

Multivariate Analysis in Weed Vegetation Research: Examples from NW Balkans

Urban Šilc[1], Sava Vrbničanin[2]

[1]Institute of Biology, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, 1000 Ljubljana, Novi trg 2, Slovenia
[2]Faculty of Agriculture, 11080 Belgrade, Nemanjina 6, Serbia

SUMMARY

Studies of weed vegetation by using Braun-Blanquet's method have been traditional in the north-western Balkans but only studies based on synoptic tables of weed vegetation have so far been conducted in the area of the former Yugoslavia (Kojić, 1975). Over the past two years, all available relevés of weed vegetation (4258) in the area have been collected in a database, which we have analysed using the latest methods of multivariate analysis in different statistical softwares, such as JUICE, CANOCO, PC-ORD, etc.

The results of classification, indirect and direct ordination and application of indicator values (Pignatti, Raunkier forms) in the analysis of weed vegetation are shown on examples from the former Yugoslavia, and will be compared to results acquired under the conventional approach of Braun-Blanquet's scool. It will show the advantages of the new approach in studies and interpretation of weed or any other type of vegetation.
Keywords: Weed vegetation; Multivariante analysis; Numerical classification; Ordination

INTRODUCTION

Research of weed vegetation commonly focuses on the influence of numerous different factors (soil, climate, crop type, crop management, weed management) on biotic data, and the use of multivariate statistics is therefore welcome, even necessary, in our search for the right answers to complex questions.

Multivariate statistical methods are powerful tools for investigating and summarizing underlying trends in complex data structures (Legendre and Legendre, 1998). Criticism has been voiced over the use of these methods to manipulate results in order to fit the user's assumptions or to reveal what is already known. Moreover, multivariate methods (a) select the most important factors from multiple factors imagined or hypothesized, (b) separate strong, important patterns from weak ones, (c) reveal unforeseen patterns and suggest unforeseen processes (McCune and Bruce, 2002).

Multivariate statistical methods in weed vegetation studies are not a novelty, there have been many review papers dealing with their application in weed science (Streibig, 1979; Post, 1988; Kenkel et al., 2002). A great number of multivariate statistical methods have been used in recent years but there have also been instances of their uncritical use according to "fashion" in any scientific discipline in which the user is active. In the computer era, several appropriate softwares (e.g. SYN-TAX, CANOCO, PC-ORD) have been designed to make these methods more accessible.

The aim of our paper is to summarize some of the multivariate methods most frequently used in weed science,, i.e. in research of weed vegetation, and to highlight examples from weed vegetation datasets from the north-western

Balkans (former Yugoslavia). For more statistical details and in-depth information, the readers are referred to other papers and books specializing in statistics.

PLANT COMMUNITY CLASSIFICATION

The Braun-Blanquet approach in plant community classification

The Braun-Blanquet plant community classification, also termed 'syntaxonomical system', is hierarchical and resembles the taxonomy system. The association is the central community type; similar associations are united in alliances, alliances in orders, and orders in classes. The association may be differentiated into sub-associations (Braun-Blanquet, 1964). Each syntaxon is defined by a characteristic species combination, a group of diagnostic taxa that may include character taxa, differential taxa and companions. The degree of differentiation within the taxon is usually the species. Character species are ideally limited to one syntaxon "A" – but need not occur in many of the relevés assigned to "A". Differential species occur in "A" but not in syntaxa within the next higher rank, while again they may occur in other syntaxa. Companions occur in most relevés of "A" but also in other syntaxa (Braun-Blanquet, 1964).

Large-scale approaches in plant community classification

European vegetation studies have recently resulted in a survey of 928 phytosociological alliances in 80 classes, each with an English name referring to the physiognomy, type of habitat and distribution area (Rodwell et al., 2002). The classes are included into 15 formations. Also, similar studies have been made in North America. American vegetation scientists define the association as a physiognomically uniform group of vegetation stands sharing one or more diagnostic elements**:** differential species, highly frequent species, indicator or character species. The alliance is a physiognomically uniform group of associations sharing one or more diagnostic species which are found in most vegetation strata. The survey lists 4657 associations grouped into 1522 alliances (van der Maarel, 2005).

Classification and ordination as complementary approaches in phytosociology

Both the American and European phytocoenological schools consider the alliance a central unit. In complementary phytosociological approaches to classification and ordination the hypothesis was that with increasing environmental complexity classification becomes less and ordination becomes more effective (van der Maarel, 2005). It has been recommended to apply both techniques while approaching syntaxonomy on the alliance/order level (van der Maarel, 1979).

QUANTITATIVE METHODS IN VEGETATION STUDIES

Numerical Classification

Numerical classification, also known as cluster analysis (Lepš and Šmilauer (2003) use this term only for agglomerative classifications), is used to define species assemblages and plant community types (Kent, 2006). It arranges datasets into hierarchical systems and it resembles the system of the Braun-Blanquet school.

There are many different methods of classification that differ regarding the starting point of classification. When they start from the bottom by grouping two most similar objects (e.g. relevés) they are called agglomerative methods, and if they start from the top by dividing datasets into two groups and so forth they are called divisive methods.

Ward's method is probably the most widely employed agglomerative method, proved to be effective, useful and one of the few space-conserving methods. A major flaw of the method is its incompatibility with Sörensen distance. The use of flexible-beta linkage (with β=-0.25) is recommended for getting similar results but without this drawback. Also transformation of data is used to improve the results.

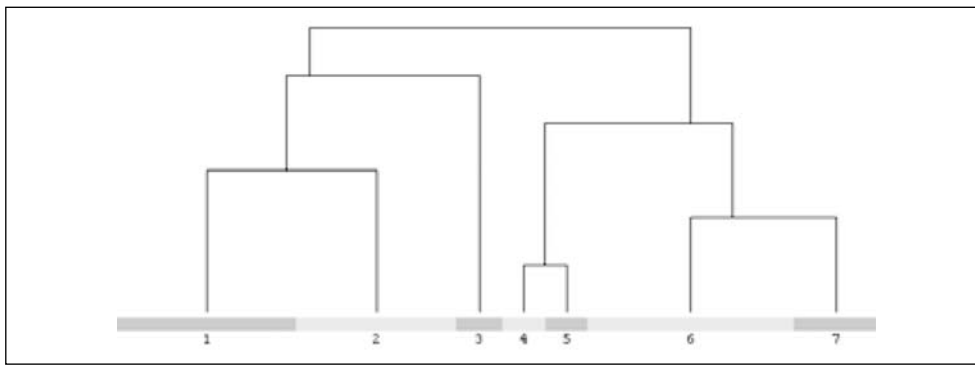

Figure 1. Dendrogram of 2426 samples of weed vegetation from NW Balkans grouped into 7 major alliances (Šilc et al., in press)
Slika 1. Dendogram za 2426 uzoraka korovske vegetacije na severozapadnom Balkanu grupisanih u 7 osnovnih sveza (Šilc i sar., u štampi)

With classification we tried to confirm (more objectively) the syntaxonomical scheme at the level of phytosociological alliance made in traditional way (Kojić, 1975; Ajder, 1992, 1997; Kojić and Vrbničanin, 1998; Kojić et al., 1996; Vrbničanin et al., 1998). All earlier subjectively made classifications have been confirmed with minor corrections for the *Galeopsion* alliance (Šilc et al., in press). On the other hand, a similar analysis did not confirm all syntaxa from the Czech Republic (Lososová et al., 2006).

The other very common divisive method used is TWINSPAN (Hill et al., 1975; Hill, 1979; Kent and Coker, 1992), which is in free use but has been criticized in recent years by McCune and Grace (2002), who do not recommend it to be used in ecology at all, except in special cases in which the dataset has a simple, one-dimensional underlying structure. But such cases do occur; we have used this method in classifying vernal weed vegetation in NW Balkans with only one gradient that was previously known (Šilc et al., in press).

Despite criticism, the method has undergone some development. Tichý et al. (2007) modified the hierarchy of divisions in the final classification tree. Divisive classification with these stopping rules tends to avoid the imposed divisions of homogeneous clusters at the higher levels of classification hierarchy.

Ordination

Ordination is even more used in ecology and there is also a wider span of methods used. Ordination is used in two ways: in a descriptive way to make an exploratory analysis of underlying data structure (e.g. species composition) or for statistical hypothesis testing as a predictive approach. Alternative terminology is the indirect and direct gradient analysis.

Indirect gradient analysis

In case our dataset consists only of species data we should use an indirect gradient analysis. On the other hand, this type of ordination is also used for exploratory analysis of data structure to find outliers or nonlinear trends (Kenkel et al., 2002) that would lead to more appropriate data transformation, dataset "cleaning" and multivariate approach.

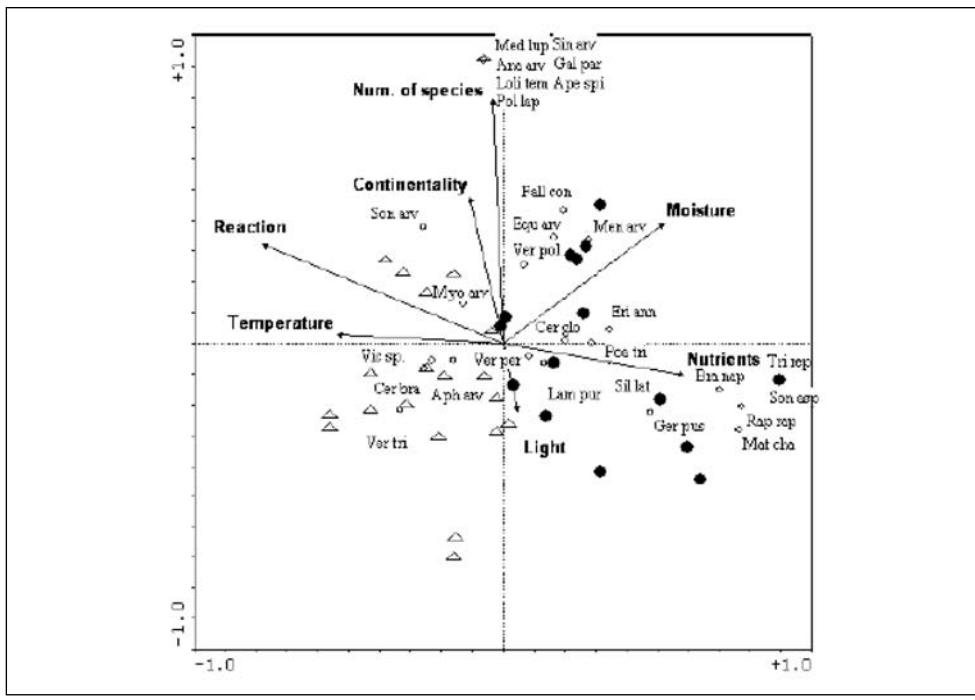

Figure 2. Ordination diagram (PCA) showing differences in species composition between 1939 (triangle) and 2002 (circle)
Slika 2. Dijagram ordinacije (PCA) koji pokazuje razlike u sastavu vrsta za 1939. (trougao) i 2002. (kružić) godinu

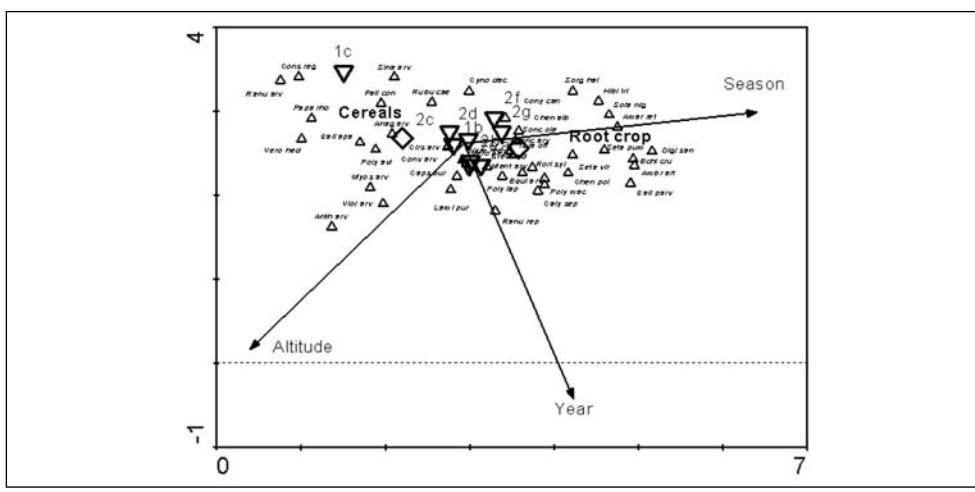

Figure 3. DCA ordination diagram showing an overall pattern of floristic composition in a large dataset from NW Balkans (Šilc et al., submitted).
Slika 3. Dijagram ordinacije DCA koji pokazuje širi obrazac floističkog sastava u okviru velike baze podataka za severozapadni Balkan (Šilc i sar., u štampi).

Numerous ordination techniques are available: Principal Components Analysis (PCA), Principal Coordinates Analysis (PCoA), Detrended Canonical Analysis (DCA) or Non-metric Multidimensional Scaling (NMDS).

Principal Components Analysis (PCA) was used to reveal temporal changes in floristic composition between 1939 and 2002, with Ellenberg indicator values passively projected (Šilc and Čarni, 2005).

But in weed vegetation analysis the DCA is probably most widely used. We use it as a first analysis of gradients in species composition. The DCA analysis of weed vegetation from the Balkans reveals a large gradient (Šilc et al., submitted) and it is similar for Central Europe (Lososová et al., 2004). This method, although indirect, also allows environmental factors to be passively projected onto the ordination diagram and to reveal the overall variation patterns in weed species composition.

Direct gradient analysis

In contrast to indirect methods, samples are positioned here with regard to the environmental factors measured in these samples. Direct gradient analysis has some advantages over indirect: it is objective, statistical p-value has its meaning and the analysis must be planned *a priori*. But it is the motive of the research that defines the choice of the method.

The most frequently used direct methods are: CCA (Canonical Correspondence Analysis) and RDA (Reliability Discriminate Analysis). Partial analyses are another advantage of these methods. They allow to determine if one set of environmental factors explains any variation in floristic composition that is not explained by another set of environmental variables.

An example of the use of CCA analysis and particularly of partial CCA was made on a dataset from the former Yugoslavia (Šilc et al., submitted). We used a set of variables (season, year, crop, phytogeographical region) that were predicted to be most influential on the floristic composition and than performed several partial CCA analyses.

What confirms the usefulness of ordination methods are the results obtained from similar researches in Slovenia (Šilc, 2008) and the Czech and Slovak Republics (Lososová et al., 2004). The results were unexpectedly different as other variables had influenced weed vegetation composition more and in a different way than they did in the Balkans. This unforeseen pattern could be only revealed by ordination methods.

Table 1. The results of partitioning of the explainable variation of species composition among five groups of variables. Gross effect is the total variation of the particular variable, shared with other variables. Net effect is the explained variation by a particular variable not shared with others. F-values are results of a permutation test of partial CCAs.
Tabela 1. Rezultati razlučivanja objašnjive varijacije u sastavu vrsta na osnovu pet grupa promenljivih. Ukupan efekat je zbir promena pojedinačnih promenljivih vrednosti u sadejstvu sa drugim promenljivim vrednostima. Neto efekat je promena objašnjena pojedinačnom promenljivom bez sadejstva sa drugim. F-vrednost je rezultat testa permutacije parcijalnih CCA.

	Gross effects	Net effects	Permutation test: F- value	p<0.01
Phytogeography	0.837	0.730	15.401	**
Crop	0.310	0.199	13.943	**
Year	0.165	0.116	8.144	**
Season	0.162	0.092	6.434	**
Altitude	0.150	0.087	6.098	**
Crop & Season	-	0.046		
Phytogeography & Year	-	0.039		
Phytogeography & Altitude	-	0.038		
Phytogeography & Crop	-	0.024		
Crop & Year	-	0.012		
All	1.419	1.419		

An excellent example of the use of weed vegetation datasets has been made by Hallgren et al. (1999), who divided their dataset into two parts. The first part was used for exploratory analysis, while the other was used for hypothesis testing. They used DCA and CCA methods with different options.

CONCLUSION

Multivariate techniques have proved to be a useful tool for weed scientists in general, not only for weed vegetation researchers as the selected examples might suggest. But researchers need to bear in mind not only the advantages of multivariate analysis, but also the theoretical background and limitations of these methods.

ACKNOWLEDGEMENT

The article is the result of a bilateral project between the Republic of Slovenia and Republic of Serbia "Weed flora and vegetation of arable land and changes in agroecosystems" No. 337-00-167/2006-01/24.

REFERENCES

Ajder, S.: The influence of some herbicides on phytocoenological changes of the corn crop weed community. Review of Research Work at the Faculty of Agriculture, 37, 1, 37-47, 1992.

Ajder, S.: Phytocoenological and ecophysiological change in weed vegetation of small grains in dependence on the altitude, Review of Researche Work at the Faculty of Agriculture, 42, 2, 39-62, 1997.

Braun-Blanquet, J.: Pflanzensoziologie. 3. Auflage. Springer-Verlag, Wien, 1964.

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

Hill, M.O.: TWINSPAN- a FORTRAN program for arranging multivariate data in an ordered two way table by classification of the individuals and the attributes. NY: Cornell University, Department of Ecology and Systematic, 1979.

Hill, M.O., Bunce, R.G.H., Shaw, M.W.: Indicator species analysis, a division polythetic method of classification, and its application to a survey of native pinewoods in Scotland . Journal of Ecology, 63, 597-613, 1975.

Kenkel, N.C., Derksen, D.A., Thomas, A.G., Watson, P.R.: Multivariate analysis in weed science research. Weed Science, 50, 281-292, 2002.

Kent, M., Coker, P.: Vegetation description and analysis: a practical approach. Belhaven Press, London, 1992.

Kent, M.: Numerical classification and ordination methods in biogeography. Progress in Physical Geography, 30, 3, 399-408, 2006.

Kojić, M., Ajder, S., Dajić, Z.: Syntaxonomical division of the weed vegetation in the region of Serbia. X Colloque International sur la biologie des mauvaises herbes, Dijon, France, 186, 1996.

Kojić, M., Vrbničanin, S.: Pregled korovske flore Srbije – Kategorije korova i njihove specifičnosti, biološke i ekološke karakteristike, taksonomska analiza, biodiverzitet, fitocenološka pripadnost. Zbornik rezimea IV jugoslovenskog kongresa o zaštiti bilja, Vrnjačka Banja, 99, 1998.

Legendre, P., Legendre, L.: Numerical Ecology. 2nd edition. Elsevier, Amsterdam, The Netherlands, 1998.

Lososová, Z., Chytrý, M., Cimalová, Š., Kropáč, Z., Otýpková, Z., Pyšek, P., Tichý, L.: Weed vegetation of arable land in Central Europe: Gradients of diversity and species composition. Journal of Vegetation Science, 15, 415-422, 2004.

Lososová, Z., Chytrý, M., Cimalová, Š., Otýpková, Z., Pyšek, P., Tichý, L.: Classification of weed vegetation of arable land in Czech Republic and Slovakia. Folia Geobotanica, 41, 259-273, 2006.

McCune, B., Grace, J.B.: Analysis of Ecological Communities. MjM Software Design, Gleden Beach, Oregon, USA, 2002.

Post, B.J.: Multivariate analysis in weed science. Weed Research, 28, 425-430, 1988.

Rodwell, J.S., Schaminee, J.H.J., Mucina, L., Pignatti, S., Dring, J., Moss, D.: The diversity of European vegetation. An overview of phytosociological alliances and their relationships to EUNIS habitats. Report EC/LNV 2002/054, Wageningen, 2002.

Streibig, J.C.: Numerical methods illustrating the phytosociology of crops in relation to weed flora. Journal of Applied Ecology, 16, 577-587, 1979.

Šilc, U.: Diversity of weed vegetation on arable land in Slovenia. Journal of Plant Diseases and Protection, Sp. Issue 21, 351-356, 2008.

Šilc, U., Čarni, A.: Changes in weed vegetation on extensively managed fields of central Slovenia between 1939 and 2002. Biologia, 60, 4, 409-416, 2005.

Šilc, U., Vrbničanin S., Božić, D., Čarni A., Dajić Stevanović, Z.: Classification of weed vegetation in northwestern Balkans. Phytocoenologia, 38, (in press).

Šilc, U., Vrbničanin, S., Božić, D., Čarni, A., Dajić Stevanović, Z.: Weed vegetation in northwestern Balkans: diversity and differentiation. Weed Research, (submitted).

***Tichý, L., Roleček, J., Zelený, D., Chytrý, M.*:** New stopping rules for TWINSPAN. European Vegetation Survey, 16th Workshop, Rome, Italy, 2007.

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

***van der Maarel, E. (ed.)*:** Vegetation ecology, Blackwell Science Ltd, 2005.

***Vrbničanin, S., Kojić, M., Karadžić, B.*:** New weed association determined in southwestern weed community. Review of Research Work at the Faculty of Agriculture, 37, 1, 37-47, 1998.

Primena metoda multivarijantne analize u proučavanju korovske vegetacije na primeru severozapadnog Balkana

REZIME

Korovska vegetacija u Europi je najviše proučena klasičnim floristično-fitocenološkim metodama i to uglavnom na regionalnom nivou. Međutim, poslednjih desetak godina sve više se upotrebljavaju veći setovi podataka (velika područja), koji su donekle uticali na drugačiji pristup tumačenju faktora koji utiču na floristički sastav i diferenciranost korovske vegetacije. U vezi s tim, analiza i interpretacija korovske vegetacije primenom multivarijacionih metoda je postala pravi trend u ovakvoj vrsti istraživanja.

Proučavanje korovske vegetacije Braun-Blanquetovom metodom ima bogatu tradiciju na tlu severozapadnog Balkana, ali do sada su uglavnom rađene samo studije iz pojedinih područja kao sinoptičke tablice pregleda korovske vegetacije za područje bivše Jugoslavije (Kojić, 1975). U zadnje dve godine u bazi podataka sakupljeni su svi dostupni vegetacijski snimci (4258 sastojina) korovske vegetacije sa tog području, koje smo analizirali koristeći najsavremenije metode multivarijantne analize u različitim statističkim programima kao što su JUICE, CANOCO, PC-ORD itd.

Na praktičnim primerima u radu će biti prikazani rezultati klasifikacije, indirektne i direktne ordinacije i primene indikacionih vrednosti (Pignatti, Raunkierovi oblici) u analizi korovske vegetacije razvijene na prostorima bivše Jugoslavije i kao takvi biće upoređeni sa rezultatima pregleda korovske vegetacije rađene klasičnim principima Braun-Blanquetove škole. Dakle, dobiće se uvid u prednosti novog pristupa u proučavanju i interpretaciji korovske, odnosno bilo koje druge vegetacije.

Multivarijantne metode omogućavaju dobijanje informacija iz više-dimenzionalnog sveta, kao što je priroda, a naročito dinamična korovska vegetacija, a koje nisu direktno vidljive. Klasifikacija korovskih/biljnih zajednica ovim sistemom omogućava objektivnije grupisanje velikog broja podataka, za razliku od starih principa Braun-Blanquetove škole gde je to rađeno intuitivno, tj. na osnovu iskustava istraživača. Ordinacionim tehnikama mogu se koristiti i merljivi parametri kao što su tip useva, godina, sezona, nadmorska visina itd., a primenom parcijanih ordinacija uticaj pojedih varijabli i nivo njihove statističke značajnosti u razvoju određenog tipa vegetacije.

Ključne reči: Korovska vegetacija; multivarijantna analiza; numerička klasifikacija; ordinacija

Received July 7, 2008.
Accepted July 10, 2008.