

Groups Analysis Of Spiders Species By Preferences For Tirana Habitat Area

Markela Muca

University of Tirana, Faculty of Natural Sciences,
Department of Applied Mathematics, Tirana, Albania

Blerina Vrenozi

Tirana University, Faculty of Natural Sciences,
Research Centre of Flora and Fauna, Tirana, Albania

doi: 10.19044/esj.2016.v12n6p356 [URL:http://dx.doi.org/10.19044/esj.2016.v12n6p356](http://dx.doi.org/10.19044/esj.2016.v12n6p356)

Abstract

In this paper, we presented the cluster analysis of ground-dwelling spiders. The scientific material was collected from different areas of the Tirana district, using traps Pitt. Preferences per habitat have shown that areas of the Veterinary Research Institute and the mount of Dajti have a larger number of species that are abundant in various habitats. Also, depending on the areas, it can be classified into three major groups: The Vora hills, Tirana city, and the Dajti Mountain. The analysis is based on a sample which consists of 123 ground-dwelling spiders scattered mostly in seven areas. The results of the analysis have shown the spider distribution in at least one of the areas included in the study. In addition, it also shows the dominant species for each area. All analysis is performed using SPSS statistics 20.

Keywords: Cluster analysis, hierarchical method, habitat, Balkan fauna, ground-dwelling spiders

Introduction

Graphical presentation of data matrix helps to detect groups. Graphical displays of multidimensional data have remained a problem till today. Therefore, we need techniques for graphical representation of data in 2-3 dimensional space. The histogram is often a useful first step in the search for detecting group's modes in data, particularly of course, if the data are univariate. More especially, in multimodal distributions, the mode is an indication of the existence of groups. In the case where we have two variables, we can cover the possible groups by scatter plot (Brian S. Everitt et al., 2011). With multivariate data, histograms can be constructed for each separate variable. Thus, in general, this will not be of great help in

uncovering cluster structure in the data because the marginal distribution of each variable may not reflect accurately the multivariate distribution of the complete set of variables.

Firstly, in this paper, we will present some graphical tools for the data from the dataset which consists of 742 individuals belonging 124 types of spider and 7 habitats. The description of variables in the data set is displayed in Table 1. The problem with this data is that they cannot be presented in a space of 8-dimension.

Another solution would be to implement one of the known reduction procedures like: PCA, FA (M. Muca et al., 2013), or MDS (Sascha Buchholz, 2010; Popov V. V., 2000; 2007). Thus, the new variables (known as component, factor, or stimuli) are presented graphically in a way to find out who will be determined by the arithmetic procedures to CA (Belkhir et al., 2011).

Spiders (Araneae) are a delicate group of arthropods and they represent a very important step of the food chains (Duffey E., 2005). Invertebrates, especially spiders, are widely used as indicators of the quality of the environment. This is because they are sensitive to environmental concerns and they show a negative effect of the management processes or the functioning of ecosystems (Blandin P., 1986). Spiders have a high potential as an indicator of environmental quality of ecosystems. They involve very sensitive species. They can be observed in a significant number for draw results assessment of the environmental situation, and are types that reflect the ecological effects of habitat concerns. Spiders are used as bio-indicators of environmental quality, because they are very effective in making comparisons and are used reliably for identification (Maelfait J.P., 1996).

In this study, we presented habitats with larger abundance types of spiders and in the grouping of species of at least one of the zones in the study.

Materials and Methods

In this study, there were seven areas in Tirana district which include urban habitats, suburban, rural, artificial forest, degraded forests, and preserved forest. However, these depend on the vegetation habitats and are subject to human impact. Furthermore, these areas were classified on three other bigger groups, depending on the proximity to each other and its height above sea level: The Vora hills, Tirana city and Dajti Mountain (Table 1).

Table 1. Description of the Variables in the Data Set

Variable	Description
G. Bollës	Stone Bolles, degraded forest shrubs located on the top of the Vora hills.
Rob_pseud	Artificial forest with <i>Robinia pseudaccacia</i> located on the Vora hills.
Olea_eur	Rural area with <i>Olea europaea</i> located on Vora hills.
MSHN	Museum of Natural Sciences, urban area situated in the center of Tirana.
NT	Hospital “Mother Teresa”, urban area, located on the center of Tirana.
IKV	Institute of Veterinary Research, bottom of Dajti mountain, suburban area.
M. Dajtit	Dajti mountain, natyral park with <i>Fagus sylvaticum</i> , northeast of Tirana city.

The material is assembled using pitfall traps. This technique was first developed by Hertz (Hertz M., 1927), who used the open-containers (plastic boxes) placed in the ground at the surface level. Pitfall trapping is one of the most commonly used ecological methods to catch active wandering and cursorial invertebrates (Vrenozi B. et al., 2012; Vrenozi B. & Jäger P., 2012).

The cluster analysis is a collection of techniques that aim to assign observations (e.g. people, things, events, companies) into groups (clusters). This was such that the observations within each group are similar to one another with respect to variables or attributes of interest, and the groups themselves stand apart from one another based on certain criteria (Han J. & Kamber M., 2000). Most of the results obtained by methods of analysis of groups have to do with the separation or agglomeration of the data groups. In this data group, every individual is part of a group or a single group that contains all observations (Everitt B. S., 1993).

Both methods (partition and agglomerative) known as hierarchical classification methods are based on the study of the distance matrix and the use of a specific optimization criterion of agglomerative or partition procedures of groups (Brian S. Everitt et al., 2011).

There are various measures to express (dis)similarity between pairs of objects. The Euclidean distance, the weighted Euclidean distance, and the Mahalanobis distance are some measures of distance for numerical variables. After choosing the distance or similarity measure, we need to decide which clustering algorithm to apply. There are several agglomerative procedures and they can be distinguished by the way they define the distance from a newly formed cluster to a certain object, or to other clusters in the solution. The single linkage, complete linkage, and Ward’s methods are the most popular agglomerative clustering procedures. However, each of these linkage algorithms can yield totally different results when used on the same dataset (Brian S. Everitt et al., 2011). In this paper, we will use the Ward’s methods (Ward J. H., 1963) which is one of the known hierarchical classification method. Also, *phi-square* distance is used as a similarity measure. Phi-square distance is used in count data and it is calculated as chi-square/total

(firstly, contingency table 2xn is calculated, and is built by all pairs of individual).

The data set has an information for n=742 individuals which belongs to 124 types of spider and seven habitats. The most dominant species were *Ozyptila confluens* (n=81), *Trochosa hispanica* (58), *Inermocoelotes falciger* (n=50), *Pardosa proxima* (n=46), *Histoipona pr. vignai* (n=40), *Hogna radiata* (n=36), and *Alopecosa pulverulenta* (n=33). Also, it distinguished the other groups of species with a smaller number of individuals; for example, six types (n=14-20 individuals), 19 types (n=5:12), and other types are found less frequently (n=1:4). Therefore, these species are distributed in at least one habitat, where the total number of habitats is equal to seven.

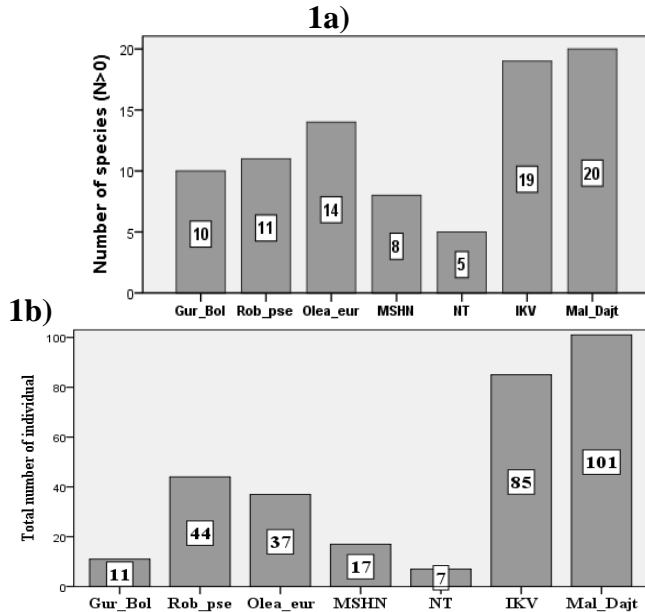
Analysis of the Types of Spiders by Dominance for Each of the Study Areas

To analyse the grouping of species that only have a habitat preference, some numerical features and the confidence interval for a mean number of individuals to type in each of the study areas was calculated (Table 2). It notes that the maximum number of species has higher values in IKV and in Mal_Dajt respectively with 46 and 40 individuals. Furthermore, it also has smaller values in Gur_Bol and NT with two individuals each.

Table 2. Some numerical value and the confidence interval for the average number of individuals to type in each of the study areas

	Number of species	Mean	Standard deviation	Standard error	95% Confidence interval for mean		Minimum	Maximum
					Lower	Upper		
					Gur_Bol	10		
Rob_pse	11	4.00	5.138	1.549	.55	7.45	1	17
Ole_eur	14	2.64	2.373	.634	1.27	4.01	1	9
MSHN	8	2.13	1.356	.479	.99	3.26	1	4
NT	5	1.40	.548	.245	.72	2.08	1	2
IKV	19	4.47	10.341	2.372	-.51	9.46	1	46
Mal_Dajt	20	5.05	9.395	2.101	.65	9.45	1	40

Graphically, it also shows the number of individuals and species that are in each area (Graphic 1a; b). As can be seen, the largest number of species and individuals were from areas IKV (19 types, 85 individuals) and Mal_Dajti (20 types, 101 individuals), while the smallest number was NT (5 types, 7 individuals).

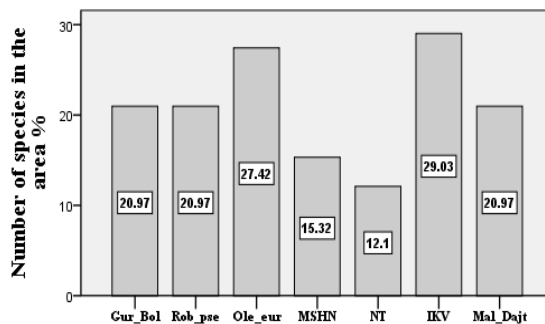


Graphic 1. The distribution of species and individuals in one area

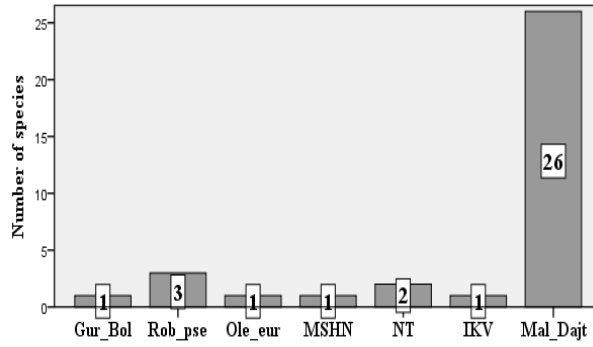
The Group of Species in at Least One of the Areas Studied

The initial community analysis of data shows the distribution of species by habitat where it was noted that species have different preferences (Graphic 1).

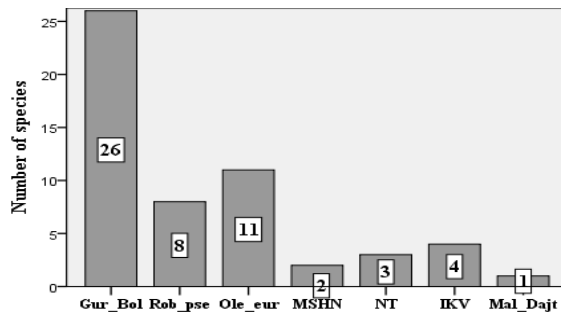
Graphic 2 shows that about 21% (26 from 124) of species are distributed at least in the area of Mal_Dajt. Also, 3 out of 26 (11.54%) were distributed in at least two areas, Mal_Dajt and Rob_pse and so on (Graphic 1a). The numbers of species that are only in Mal_Dajt comprise a total of 20 species which only has this habitat preference (Graphic 2). *Histopona prope vignai*, *Dysdera storkani* dhe *Inermocoelotes microlepidus* (n=40, 19 and 12).



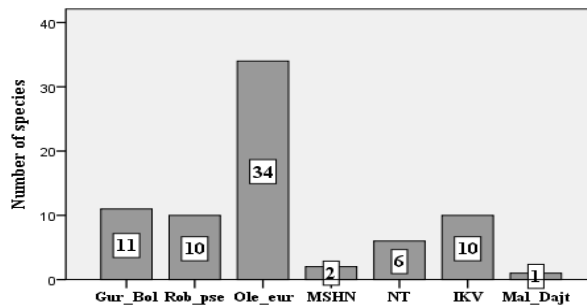
Graphic 2. The distribution of the species in at least one of the habitats



Graphic 2a. The distribution of species at least in the area of Dajti mountain



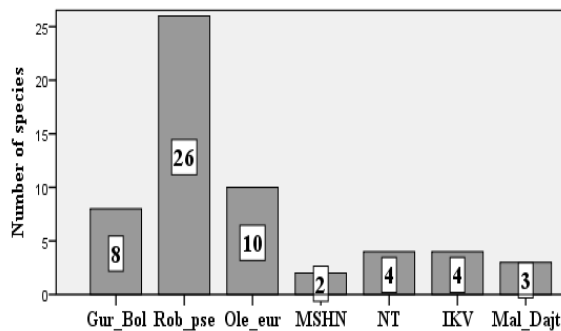
Graphic 2b). The distribution of species at least in the area of Stone Bolles



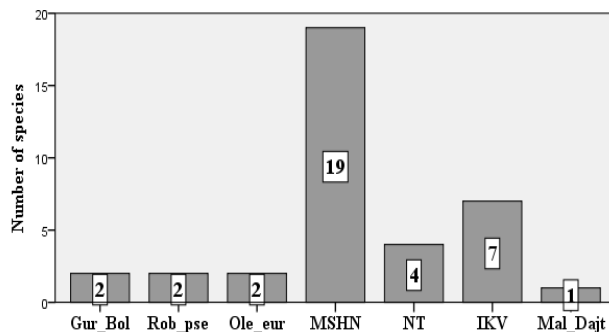
Graphic 2c). The distribution of species at least in the area of **Olea europaea**

Graphic 2, shows that about 21% (26 out of 124) of the species are distributed at least in the area of **Stone Bolles**. However, 11 out of 26 (42.31%) and 8 out of 26 (30.77%) are distributed in at least three areas, namely: Gur_Bol, Ole_eur, and Rob_pse. Thus, species that are at least in

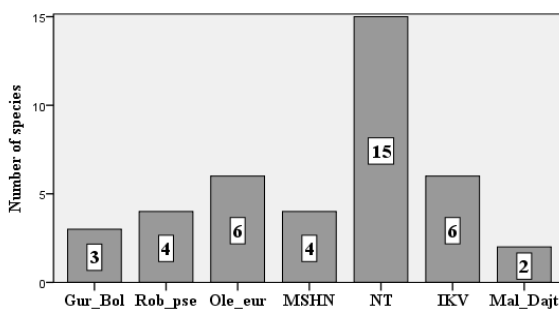
areas of Gur_Bol have the greatest preference for areas Ole_eur and Rob_pse (Graphic. 2b). *Ozyptila confluens*, *Hogna radiate*, and *Alopecosa pulverulenta* (n=16, 11 and 17) are the three most dominant species in this area. Furthermore, Graphic 2c) shows that species that are at least in rural area with *Olea europaea*, have the greatest preference for areas Gur_Bol (32.35%), Rob_pse (29.41%), and IKV (29.41%). *Ozyptila confluens*, *Hogna radiata*, *Alopecosa pulverulenta*, and *Xysticus bufo* (n=51, 18, 11, and 13) are four dominant types in this area. Graphic 2d shows that species that are at least in the area with *Robinia pseudaccacia* have the greatest preference for areas Gur_Bol (30.77%) and Ole_eur (38.46%). *Ozyptila confluens*, *Inermocoelotes falciger*, *Trachyzelotes pedestris*, and *Drassyllus villicus* (n=14, 22, 17 and 17) are four dominant types in this area. Graphic 2e shows that species that are at least in the area of the **Museum of Natyral** have the greatest preference area of IKV (30.77%). *Trochosa hispanica* (n=13) is the most dominant type in MSHN. Graphic 2f, shows that species that are at least in the area of the **Institute of Veterinary Research** have greater preference for area Ole_eur (27.28%). *Pardosa proxima*, *Trochosa hispanica*, *Inermocoelotes falciger*, *Tallusia vindobonensis*, and *Pardosa hortensis* (n=46, 29, 15, 13 and 11) are five dominant types in the area of IKV. Graphic 2g, shows species that are at least in the area of **Mother Teresa** which have greater preference for areas of Ole_eur (40%) and IKV (40%). *Trochosa hispanica* (n=6) is the most dominant species in the area of NT. However, this is the area with fewer species and a small number of individuals (Graphic 2).



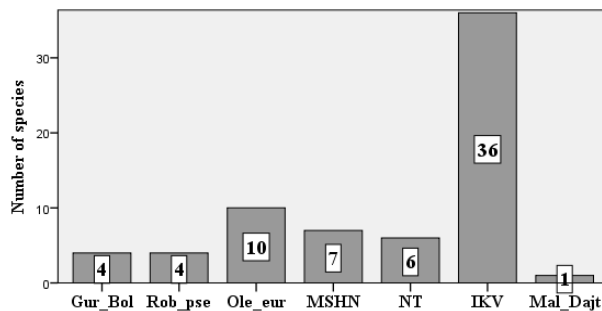
Graphic 2d). The distribution of species at least in the area of *Robinia pseudaccacia*



Graphic 2e). The distribution of species at least in the area of Museum of Natural Sciences kenceve



Graphic 2g). The distribution of species at least in the area of "Nënë Terezës"



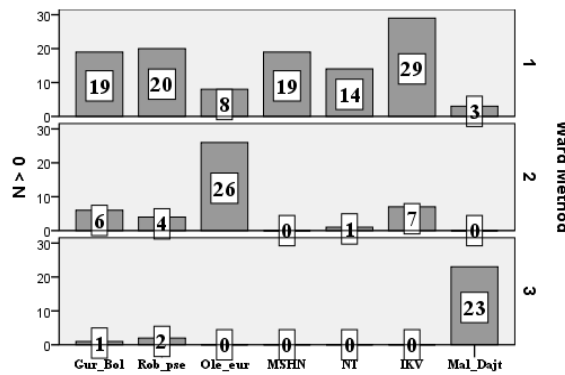
Graphic 2f). The distribution of species at least in the area of Institute of Veterinary Research

Looking at the graphs above, we can state the grouping of species by habitat. For example, we can think of grouping in a separate group those types of spiders that are found only in areas of Dajti (Graphic 2a).

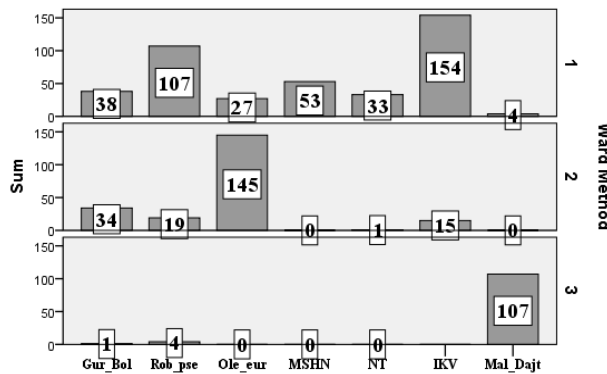
Results from Cluster Analysis

The analysis of groups can be used in the set of data in order to group types of spiders under a new structure of habitats. Similarity measure between the two groups uses the *phi-square* distance, while the measure of distance within groups uses the *Ward’s method*.

Subsequently, the analysis showed that the species of spiders can be classified into three groups (Graphic 2a; 2b).



Graphic 3a. Bar diagram by habitats (species)



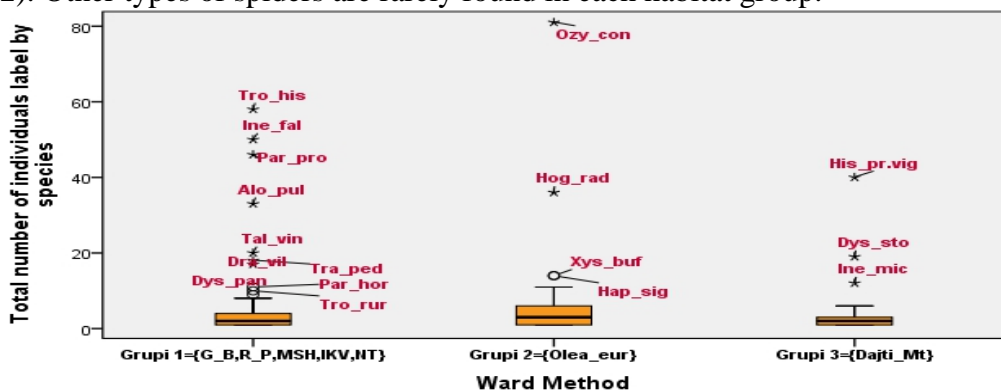
Graphic 3b. Bar diagram by habitats (individus)

Graphic 3a and 3b displayed the distribution of species and individuals of spider per habitats by groups. The grouping is done as follows:

- The first groups are those types of spiders which are located in at least two areas, and which represents the largest number of individuals for each area.
- The second groups are those types of spiders which are located mainly in a rural area with *Olea europaea*, and which have the largest number of individuals.
- The third groups are those types of spiders which are located almost exclusively in Mount Dajti, with the exception of three types which are located in Vora hills (Stone Bolles and *Robinia pseudaccacia*).

Localization of Spider Species by the Abundance for Each of the Habitats

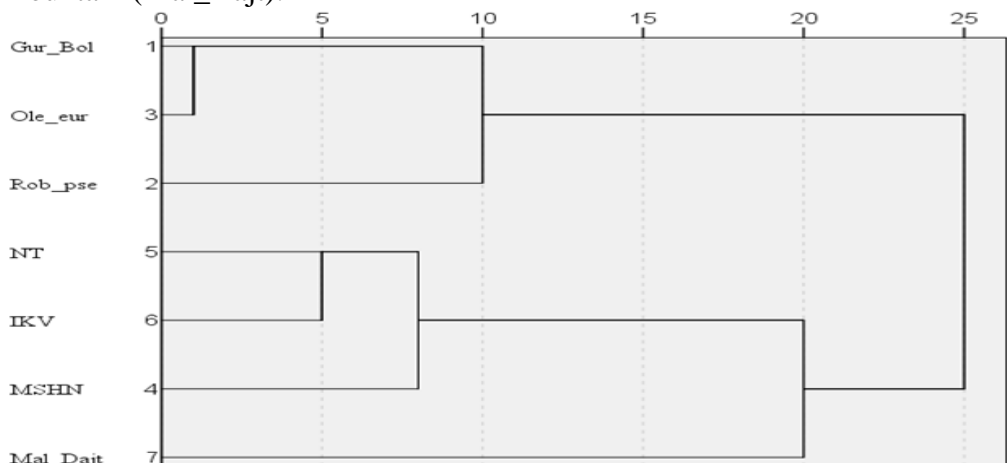
The analysis of groups shows that spider species in the database community are found at different density mostly in 7 habitats. Also, they are found at different frequency mostly in 7 habitats. In Table 3a, 3b, 3c, 3d and 3e (see, Appendix), it is seen in most species of spiders in the group 1. They are found from 1 or 2 spiders. Also, it shows that there is little spider species which are found wholesale only in group 1. So, we can say that the group 1 consists of rare species of spiders distributed mostly in 5 habitats (Gur_ Bol, Rob_pse, IKV, MSHN, and NT). Also, they are distributed by a small number of species that are abundant in these habitat. Table 3f shows that Group 2 consists of rare species of spiders mainly distributed in area *Olea_eur* and by a small number of species that are abundant in this area. Table 3g shows that Group 3 consists of rare species of spiders distributed only in the Mal_Dajt and by a small number of species that are abundant in this area. Graphic 4 shows 10 habitat types found wholesale in Group 1 in at least one of the areas. 4 types of habitats in Group 2 abound area *Olea europaea*, and 3 types of habitats in group 3 are abundant in the area Dajtit (*Dysderocrates storkani* = 19, *Histopona pr. vignai* = 40, and *Inermocoelotes microlepidus* = 12). Other types of spiders are rarely found in each habitat group.



Graphic 4. Box plot diagram for the total number of individuals in each groups (habitats)

Classification of Habitats

Using the Ward's methods (see Graphic 5), the habitat grouping according to their similarity in three areas depending on the preferences of the species can be realized. These areas include: (1) Vora hills (Gur_Bol, Ole_eur, Rob_pse); (2) Tirana city (NT, IKV, MSHN); and (3) Dajti mountain (Mal_Dajt).



Graphic 5. Dendrogram of the classification of habitats using Ward's method

References:

- Belkhiri et al. (2011). A multivariate Statistical Analysis of Groundwater Chemistry Data. *Int. J. Environ. Res.*, 5(2):537-544, Spring 2011. ISSN: 1735-6865
- Blandin P. (1986). Bioindicateurs et diagnostic des systemes écologiques. – *Bulletin d'écologie*, 17: 215 – 317.
- Brian et al. (2011). *Cluster Analysis 5th Edition*. King's College London, UK. ISBN 978-0-470-74991-3 (hardback). QA278.E9 2011. 519.5'3–dc22
- Duffey E. (2005). Regional variation of habitat tolerance by some European spiders (Araneae). – *Arachnologische Mitteilungen*, 29: 1 – 70.
- Everitt B. S. (1993). "Cluster Analysis", 3rd Edition, Edward Arnold Publishers.
- Maelfait J. P (1996). Spiders as bioindicators. In: Van Straalen N. M., D. M. Krivolutsky (eds): *Bioindicator systems for soil pollution*. – Dordrecht, Kluëer Academic Publishers: 165 – 178.
- Han J. & Kamber M (2000). "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, August.
- Hertz (1927). Huomioita petokuoriaisten olinpaikoista. – *Luonnon Ystävä*, 31: 218–222.
- Vrenosi B., Haxhiu I. & M. Koni (2012). Pitfall traps – A quantitative method for the evaluation of invertebrate distribution in Tirana district. *Journal of International Environmental Application & Science*, 7: 423-426.

Vrenosi B. & P. Jäger (2012). A faunistic study on ground-dwelling spiders (Araneae) in the Tirana district, Albania. – Arachnologische Mitteilungen 44: 81-87. Germany.

M. Muca et al. (2013). “Principal Components and the Maximum Likelihood Methods as tools to Analyze Large Data with a Psychological Testing Example”. European Scientific Journal. Vol. 9. No.20. June 2013. ISSN: 1857-7431 (on line), ISSN: 1857-7881 (Print) <http://eujournal.org/index.php/esj>.

Popov, V. V. (2000). Epigeobiont Animal Assemblages from Two Landscapes of the Bulgarian Black Sea Coast: Relationship to Environmental Gradients, Assemblage Structure and Biodiversity. III. Small Mammals (Mammalia: Insectivora, Rodentia). Acta zool. bulg., 52 (3): 79-96.

Popov, V. V. (2007). Biogeographical and ecological spatial patterns of terrestrial mammals in Bulgaria. In: Fet, V., Popov, A. (eds.). Ecology and Biogeography of Bulgaria. Monographiae Biologicae, Vol. 82: 9 – 38. Springer. ISBN: 978-1-4020-4417-5.

Sascha Buchholz (2010). Ground spider assemblages as indicators for habitat structure in inland sand ecosystems. Springer Science + Business Media B.V. 2010.

Ward, J. H. (1963). Hierarchical groupings to optimize an objective function. Journal of the American Statistical Association, 58, 236–244.

Appendix

Table 3. Contingency Table of variables with groups

		Guri i Bolles								Robinia pseudaccacia									
Table 3a		1	2	4	11	16	17	Table 3b		1	2	3	4	5	7	10	14	17	22
Ward's Method	count	15	3	0	0	0	1	count	9	2	0	1	2	2	1	0	2	1	
	1 % within Gur_Bol	78.9 %	100 %	0.0 %	0.0 %	0.0 %	100.0 %	1 % within Rob_pse	75.0 %	100 %	0.0 %	100 %	100 %	100 %	100 %	0.0 %	100 %	100 %	
	count	3	0	1	1	1	0	count	2	0	1	0	0	0	0	1	0	0	
	2 %within Gur_Bol	15.8 %	0.0 %	100 %	100 %	100.0 %	0.0 %	2 % within Rob_pse	16.7 %	0.0 %	50.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100 %	0.0 %	0.0 %	
	count	1	0	0	0	0	0	count	1	0	1	0	0	0	0	0	0	0	
	3 % within Gur_Bol	5.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3 % within Rob_pse	8.3 %	0.0 %	50.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	
	Total count	19	3	1	1	1	1	Total count	12	2	2	1	2	2	1	1	2	1	

Tab. 3c											Tab. 3d					Tab. 3e									
IKV											MSHN					NT									
1											1					1									
2											2					2									
3											3					3									
Total											Total					Total									
Ward's Method	count	17	3	0	3	1	1	1	1	1	1	count	7	3	5	3	1	count	6	3	1	3	1		
	% within IKV	810	100	0.0	60.	100	100	100	100	100	100	% within MSHN	100	100	100	100.	100.	% within NT	85.7	100	100	100	100		
	count	4	0	1	2	0	0	0	0	0	0	count	0	0	0	0	0	count	1	0	0	0	0		
	% within IKV	19.	0.0	100	40.	0.0	0.0	0.0	0.0	0.0	0.0	% within MSHN	0.0	0.0	0.0	0.0	0.0	% within NT	14.3	0.0	0.0	0.0	0.0		
	count	0	0	0	0	0	0	0	0	0	0	count	0	0	0	0	0	count	0	0	0	0	0		
	% within IKV	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	% within MSHN	0.0	0.0	0.0	0.0	0.0	% within NT	0.0	0.0	0.0	0.0	0.0		
Total	count	21	3	1	5	1	1	1	1	1	1	Total	count	7	3	5	3	1	Total	count	7	3	1	3	1

Tab. 3f														Tab. 3g								
Olea_eur														Mal_Dajt.								
1														1								
2														2								
3														3								
Total														Total								
Ward's Method	Count	4	2	0	0	0	0	1	0	1	0	0	0	count	2	1	0	0	0	0	0	
	% within Ole_eur	30.8	25.0	0.0	0.0	0.0	0.0	100.	0.0	100.	0.0	0.0	0.0	% within Mal_Dajt	15.4	16.7	0.0	0.0	0.0	0.0	0.0	
	Count	9	6	3	1	1	1	0	2	0	1	1	1	count	0	0	0	0	0	0	0	
	% within Ole_eur	69.2	75.0	100.	100.	100.	100.	0.0	100.	0.0	100.	100.	100.	% within Mal_Dajt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Count	0	0	0	0	0	0	0	0	0	0	0	0	count	11	5	3	1	1	1	1	
	% within Ole_eur	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	% within Mal_Dajt	84.6	83.3	100	100	100	100	100	
Total	count	13	8	3	1	1	1	1	2	1	1	1	1	Total	count	13	6	3	1	1	1	1