

Statistical methods in the context of environmental quality assessment: On the example of riparian beetles (Coleoptera, Carabidae)

Marina Kirichenko-Babko^{1*}, Yaroslav Danko², Lukasz Guz³, Dariusz Majerek⁴, and Roman Babko^{1,*}

¹Schmalhausen Institute of Zoology, Department Fauna and Systematics of Invertebrates NAS of Ukraine, 01030 Kyiv, Ukraine

²Sumy Makarenko State Pedagogical University, Ukraine

³Lublin University of Technology, Faculty of Environmental Engineering, Nadbystrzycka 40B, 20-618 Lublin, Poland

⁴Lublin University of Technology, Faculty of Fundamentals of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland

Abstract. Modern environmental research involves the use of a broad range of statistical methods and programs. In any study, the results are obtained, as a rule, using statistical calculations, beginning with the calculation of reliability of data, and ending with complex statistical methods that allow estimating trends and predicting the processes that are evaluated. However, there is a problem associated with the correct interpretation of biological results by experts in statistics and, accordingly, the correctness of the choice and application of various statistical methods by biologists. In this paper, the evaluation of the results of studies of the effect of changes in the hydrological regime of the river caused by the functioning of the dam on the structure and spatial distribution of riparian carabids was considered. On the banks of intact rivers, the structure of the riparian assemblage naturally changes depending on the quality of coastal sediments. In order to assess the changes in the structure of the riparian assemblage, a number of statistical methods were employed. It was shown that not every method will adequately interpret the results of ecological studies. The differences in obtaining adequate interpretations while using various statistical methods were understood.

1 Introduction

It is not possible to conduct modern ecological studies without the use of a rich arsenal of statistical methods that allow us to evaluate the trends of various processes and predict their development. On the one hand, there is a problem in choosing the method used to solve an issue, and on the other, it is difficult to adequately interpret the results of statistical analysis, as a tool to confirm or refute the original hypothesis. The use of statistics is especially important when solving problems in the field of nature protection, for example when it is required to evaluate the state of a natural object on the basis of biological data or to assess the impact of this activity on biological objects on the basis of data regarding the change in a natural object as a result of economic activity.

Today, however, most rivers are channelized and impacted by flow regulation in majority of countries [1, 2]. Arthropod abundance within the narrow shoreline strip provides the best indicator of potential trophic linkages between the aquatic and terrestrial systems as trophic interactions occur predominantly close to the shoreline [3]. The riparian arthropods are assumed to be particularly sensitive to hydrological and morphological

river modifications [4-10]. For instance, in the UK 6 beetles on exposed riverine sediments are listed as endangered and 7 vulnerable [7]. The previous study showed the impact of hydrological alteration on the spatial distribution of *Omophron limbatum* [11-12].

The use of some statistical techniques often applied in environmental engineering [13-15] was shown in this paper on the data obtained from the study on the 180 km section of the river Psel (Dnipro, Ukraine), which is divided into two sections: upper the dam at 165 km (Nizy) and lower the dam a distance of 10 km. We expect a priori that (1) or the grouping of the studied riverine habitats can be based on their types (bars, low or high bank, rampart) on the studied section of the Psel; (2) or the grouping of the studied riverine habitats can be determined by the influence of the hydroelectric dam on the spatial distribution of riparian ground beetles.

2 Method

2.1 Statistical analysis

Multivariate technique was used to describe the patterns of change in the spatial distribution of riparian ground beetles in relation to the channelized river sections and anthropogenic flow regulation. We used non-metric

* Corresponding author: rbabko@ukr.net

multidimensional scaling (nMDS) since it is particularly appropriate for the kinds of ecological data and provides a stress-factor which indicates the stability of the ordination [16-17]. The best technique for nMDS is described by Shepard [18] and Kruskal [19]. The goodness of fit of the ordination is expressed by the level of stress function and illustrated on the Shepard diagram. The nMDS was applied with similarity measures: Euclidean, Correlation, Jaccard, Simpson, Bray Curtis, Morisita, to assess why were the sample sites separating. Euclidean distance and Correlation coefficient are standard methods, which frequently used in many applications, including biological ones. Jaccard index and Simpson index are similarity measures based on presence-absence of species in ecosystems. Bray-Curtis measure is most popular abundance indicator used in biological and ecological applications. Morisita's overlap index is a measure of dispersion of individuals in a population. All the measures listed above are appropriate for this study and frequently used in the ecological analysis [20-24].

Furthermore, the hierarchical cluster analysis was performed. For ecological data, this analysis classifies sites, species or variables. In many branches of science, the agglomerative techniques are more frequently employed than the divisive techniques [25]. In our research, an agglomeration method – paired group technique, which defines cluster proximity to be the average pairwise proximities of all pairs of points from different clusters, was used [25]. Thus, all the sampling sites were grouped based upon their similarity of carabid beetles occurrence, using the following similarity indices: Correlation, Jaccard, Simpson, Bray-Curtis, Kulczynski, Chord, Euclidean distance, and with Ward's method. Hierarchical cluster analysis with Ward's method of agglomeration is frequently used with Euclidean metrics. Grouping individuals with this method tends to minimize the within-group variance [26-28]. We used PAST 2 [29] to perform the statistical analyses on the basis of a single data matrix that includes abundances of 21 riparian ground beetles and 38 sampling sites. The data are normalized by dividing variables by their standard deviations.

2.2 Study area and sampling methods

A section of the Psel (tributary of river Dnipro), a length of about 180 km, from Miropillya to the Petrenkove (Sumy district, Ukraine) was investigated from July to August 2008. We investigated the heterogeneity of riverine habitats: 20 sandy bars, 5 sites of low banks, 2 high banks, 6 sites of scarp and 4 sites of rampart (**Table 1**). Altogether, 38 sampling sites were placed randomly within study section upper and lower on the different distances from the hydroelectric power station (HPS) which is located in the Nizy. In order to compare the ecological effects of flow regulation and morphological alteration on responses of riparian beetles, all sampling sites in the studied section of the Psel river were separated between three subsections: the first – upstream to Sumy city (19 sites between 166 and 52 km), the second – above Sumy city and up to the hydroelectric dam (8 sites from 41 to 1 km), and the third –

downstream from the hydroelectric dam (11 sites from 1 to 9.7 km).

Ground beetles were sampled using plastic pitfall traps that were arranged parallel with the waterline and filled with-drop of detergent. Pitfalls were emptied after 1 to 5 days. A total of 611 traps were placed in the heterogeneous banks. The ground beetle abundances were standardized by the number of trap-days.

Table 1. Characteristics of studied sites along riverbanks Psel and their codes with kilometres (number).

| N | Name sites | Codes |
|----------------------------|------------|-------|
| Subsection 1 st | | |
| 1. | Low bank | b-166 |
| 2. | High bank | b-165 |
| 3. | Sandbank | b-164 |
| 4. | Scarp | b-163 |
| 5. | Sandbank | b-147 |
| 6. | Low bank | b-130 |
| 7. | High bank | b-91 |
| 8. | Sandbank | b-85 |
| 9. | Sandbank | b-84 |
| 10. | Scarp | b-83 |
| 11. | Sandbank | b-82 |
| 12. | Sandbank | b-80 |
| 13. | Sandbank | b-72 |
| 14. | Sandbank | b-71 |
| 15. | Sandbank | b-59 |
| 16. | Ripal | b-55 |
| 17. | Scarp | b-54 |
| 18. | Low bank | b-53 |
| 19. | Low bank | b-52 |
| Subsection 2 nd | | |
| 20. | Rampart | b-41 |
| 21. | Sandbank | b-40 |
| 22. | Sandbank | b-39 |
| 23. | Sandbank | b-38 |
| 24. | Sandbank | b-35 |
| 25. | Sandbank | b-32 |
| 26. | Sandbank | b-31 |
| 27. | Sandbank | b-30 |
| Subsection 3 rd | | |
| 28. | Scarp | b-1 |
| 29. | Scarp | b-8.7 |
| 30. | Sandbank | b-8.8 |
| 31. | Low bank | b-8.9 |
| 32. | Rampart | b-9 |
| 33. | Sandbank | b-9.1 |
| 34. | Scarp | b-9.2 |
| 35. | Sandbank | b-9.3 |
| 36. | Rampart | b-9.4 |
| 37. | Sandbank | b-9.5 |
| 38. | Rampart | b-9.7 |

3 Results

A total of 21 ground beetles collected in the section of river Psel, representing the riparian group which is widely distributed on the banks of lowland rivers in Europe [8, 30, 31]. The matrix of these taxa sampled in 38 sites was used for both multivariate analyses.

3.1 Multidimensional scaling

Two nMDS plots demonstrate the stability of ordination (Fig. 1A, B). As it can be seen, the 2D-configurational state of species composition was considered stable for Euclidean and Correlation similar measures (stress 0.09 and 0.16, respectively). Other indices used in the nMDS ordination showed a significantly lower ordination stability: Jaccard (stress = 0.28), Simpson (0.32), Bray Curtis (0.27), Morisita (0.27).

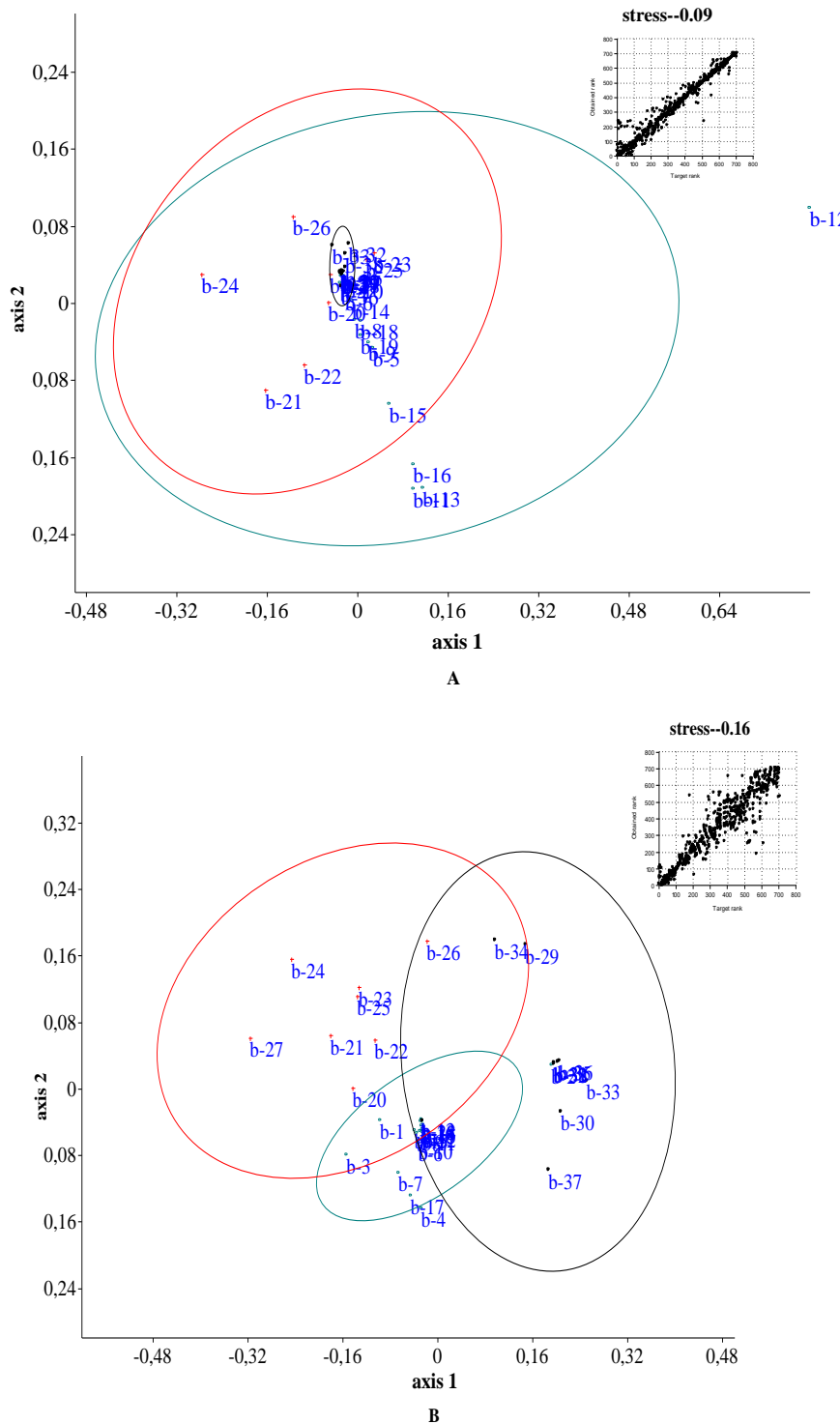


Fig. 1. Results of the nMDS ordination: (A) Euclidian and (B) Correlation of the riverine sites, based on the relative abundances of riparian carabid beetles. Codes of site names are given in Table 1. Symbols indicate of sampling sites belonging to different subsections: green oval – sites in subsection 1st, red cross – sites in subsection 2nd, black dot – sites in subsection 3.

The location is difficult to interpret (Fig. 1A). The grouping of sampling sites with using Correlation confirms the hypothesis about the impact of the hydroelectric dam on the distribution of riparian ground beetles. The highly disturbed sites in riverbanks are located on the right side (black circle) and left side (red circle) in the nMDS plot (Fig. 1B); and sites of more natural, degree of flooding are found in the centre of this plot (Fig. 1B, green circle).

The analysis of the nMDS ordination confirmed the hypothesis that the hydrological conditions changing under the influence of hydroelectric dam became the determining factor for the grouping of stations.

3.2 Hierarchical cluster analysis

The hierarchical cluster analysis supported the results of the ordination by the nMDS ordination. The Correlation (coph.corr. 0.98) and Chord distance (coph.corr. 0.99) were more adequate for interpreting our data (Fig. 2A, B) than other calculated indices. As it can be seen, both dendrograms are shown with similar results. The cluster analysis separated the 38 sampling sites into three groups corresponding to their positions in relation to the flow regulation and morphological alteration of the riverbed by HPS.

In one group, the riverine habitats stood out on the subsections which were located above the hydroelectric dam: is the II cluster includes the sites from 166 to 40 km (Fig. 2, red colour) and III cluster encompasses the sites from 40 to 30 km (Fig. 2, green colour). It should be noted that in both cases, the site N28 “scarp” is located below the dam on the 1km and included to cluster III. On the other hand, group I combined the sites which were located below dam, i.e. in banks of river reservoir (Fig. 2 A, B; subsection 3). This confirms our hypothesis that the riverine habitats in the zone of HPS influence differ significantly in terms of the structure of the riparian ground beetles assemblage.

The most common distance measure used for interval data is the Euclidean measure. It is clear that the results of the construction based on the Euclidean distance and the Ward’s method do not agree with the expectation and with the hypothesis (Fig. 3A, B). Although both clusters have a high correlation coefficient (more than 0.95), there is practically no division into separate groups. Both indices grouped the study sites without a system and, according to the results obtained, the studied section of the river looks homogeneous.

4 Conclusion

The statistical methods at a high level of reliability showed that the alterations of the hydrological regime and channel morphology are crucial for the distribution of riparian carabids. The use of statistical methods also enabled to reject one of the initial hypotheses: that the determining factor for the riparian carabid beetles is the type of shore. Consequently, the use of statistical methods without a preliminary hypothesis can lead to inadequate conclusions and complicate the interpretation of the results.

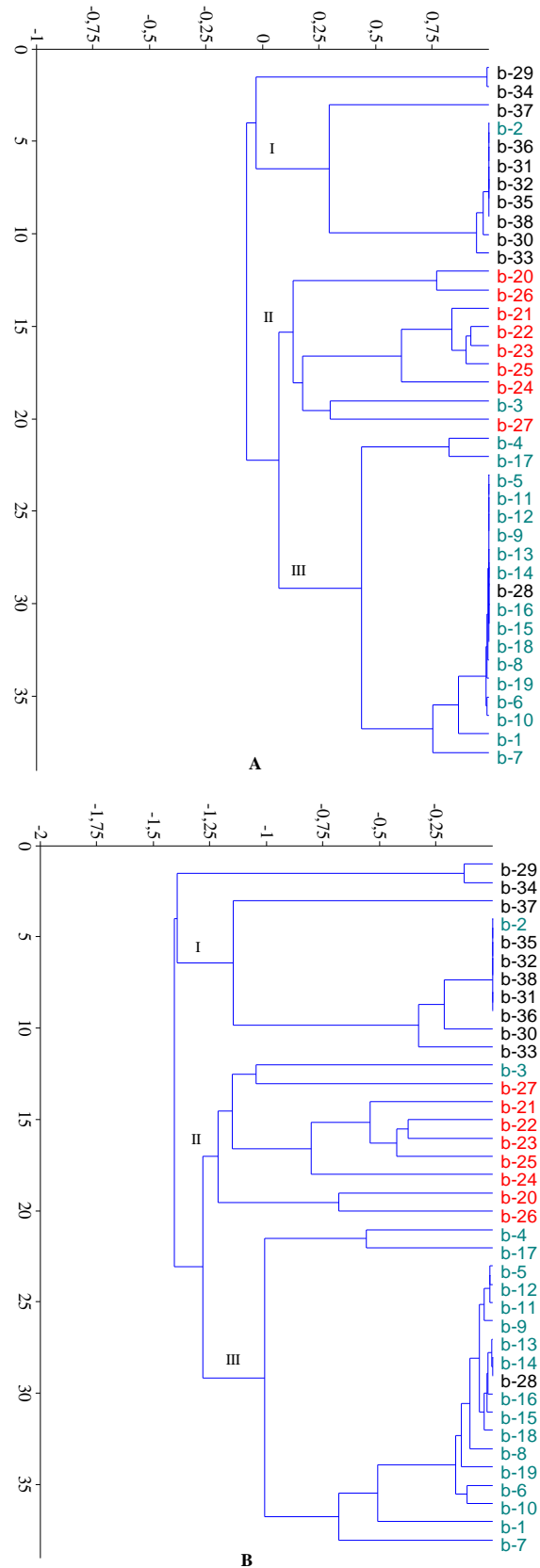


Fig. 2. Results of the similarity tree with using: (A) Correlation and (B) Chord distance with paired group linkage of 38 sampling sites, based on the relative abundances of riparian carabids species. Codes of site names are given in Table 1; sampling sites belonging to different subsections indicate by colours: green – sites in subsection 1st, red – sites in subsection 2nd, black – sites in subsection 3.

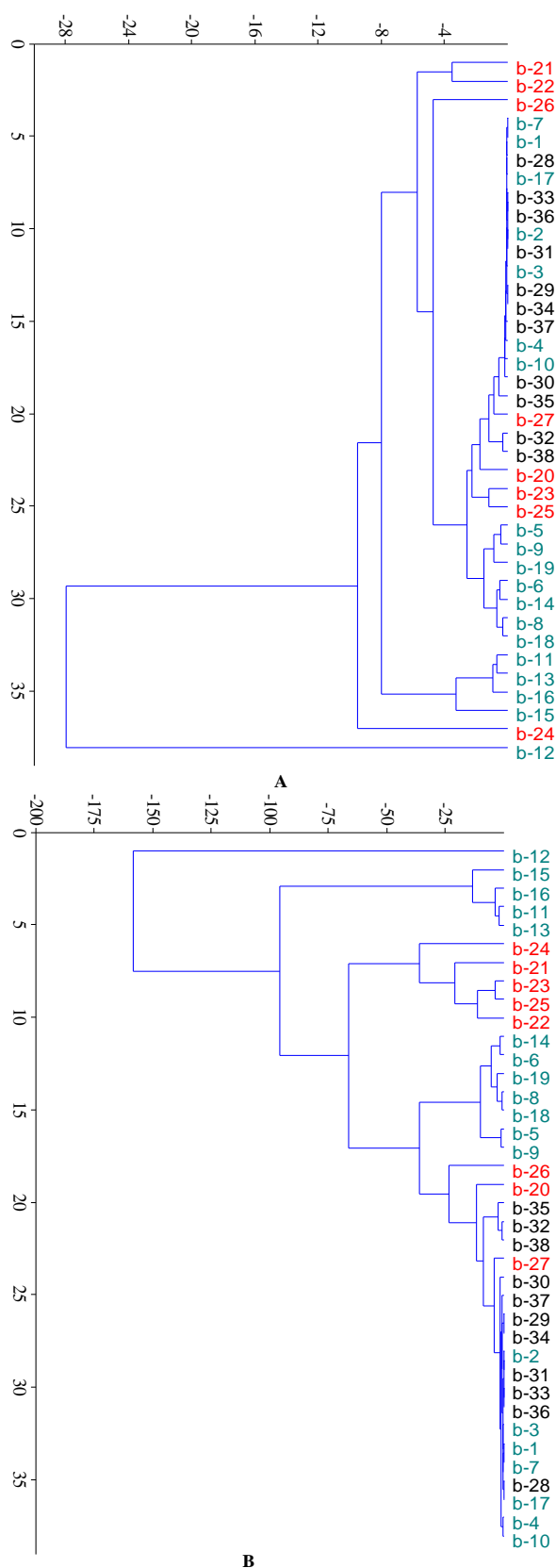


Fig. 3. Results of the similarity tree with using (A) Euclidian distance with paired group linkage and (B) Ward's method of 38 sampling sites, based on the relative abundances of riparian carabids species. Codes of site names are given in Table 1; sampling sites belonging to different subsections indicate by colours: green – sites in subsection 1st, red – sites in subsection 2nd, black – sites in subsection 3.

References

1. K. Tockner, J.A. Stanford. *Environ. Conserv.* **29**, 308–30 (2002)
2. Tockner K., Paetzold A., Karaus U., Claret C., Zettel J. Ecology of braided rivers. *Braided Rivers - IAS Special Publication* (eds G.H. Sambrook Smith, J.L. Best, C.S. Bristow & G. Petts, Blackwell, Oxford, 2006)
3. A. Paetzold, C.J. Schubert, K. Tockner. *Ecosystems* **8**, 748–59 (2005)
4. L.M. Ellis, C.S. Crawford, Jr., M.C. Molles. *Regul. Rivers: Res. Manage.* **17**, 1–20 (2001)
5. R. Manderbach, D. Hering. *Archiv Hydrob.* **152**, 583–608 (2001)
6. A. Bonn, K. Hagen, D. Wohlgemuth-von Reiche, *River Res Appl* **18**, 43–64 (2002)
7. J.P. Sadler, D. Bell, A. Fowles. *Biol. Conserv.* **118**, 41–56 (2004)
8. J. Andersen, O. Hanssen. *Biodiversity Conserv.*, **14**, 3497–3524 (2005)
9. S.R. Laeser, C.V. Baxter, K.D. Fausch. *Ecol. Res.* **20**, 646–651 (2005)
10. A. Paetzold, C. Yoshimura, K. Tockner. *J. App. Ecol.* **45**, 3, 894–903 (2008)
11. J. Günter, B. Hölscher, F. Prüßner, T. Assmann, *Mitteilungen der Deutschen Gesellschaft für Allgemeine und Angewandte Entomologie*, **14**, 517–520 (2004)
12. M. Kirichenko, R. Babko. *Teka Kom. Ochr. Kszt. Środ. Przyr. – OL PAN*, **6**, 129–137 (2009)
13. M. Chomczyńska, A. Montusiewicz, J. Malicki, G. Łagód, *Environ Eng Sci.* **26**, 2, 289–295 (2009)
14. G. Łagód, M. Chomczyńska, A. Montusiewicz, J. Malicki, A. Bieganski, *Ecol. Chem. Eng.* **16**, 3, 369–377 (2009)
15. G. Łagód, M. Chomczyńska, A. Montusiewicz, J. Malicki, D. Stransky, *Ecol. Chem. Eng.* **21**, 4, 593–604 (2014)
16. R.H.G. Jongman, C.J.F. Ter Braak, O.F.R. Van Tongeren, *Data analysis in community and landscape ecology* (Cambridge Univ. Press, 1995)
17. M.L. Beals. *Oecologia* **150**, 484–495 (2006)
18. R.N. Shepard. *Psychometrica*, **27**, 219–246 (1962)
19. J.B. Kruskal. *Psychometrica* **29**, 1–27 (1964)
20. Y. Cao, W.P. Williams, A.W. Bark. *Water Environ Res.* **69**, 95–106 (1997)
21. M.D. Mountford. *An index of similarity and its application to classification problems* (Progress in Soil Zoology, 43–50. Butterworths, 1962)
22. D.P. Faith, P.R. Minchin, L. Belbin, *Vegetatio* **69**, 57–68 (1987)
23. C.M. Hwang, M.S. Yang, W.L. Hung. *Int J Intell Syst.* **33**, 8, 1672–88 (2018)
24. J. Wo, X.X. Mou, B.D. Xu, Y. Xue, C.L. Zhang, Y.P. Ren, *Chin. J. Appl. Ecol.* **29**, 1, 285–92 (2018)

25. J.A. Ludwig, J.F. Reynolds. *Statistical Ecology: a primer on methods and computing* (1988)
26. L. Kaufman, P.J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, New York, 1990)
27. J.H. Jr, J. Ward, J Am Stat Assoc. **58**, 236–244 (1962)
28. R.S. King, *Cluster Analysis and Data Mining: An Introduction* (Dulles, Virginia: Mercury Learning & Information, 2015)
29. Ø. Hammer, D.A.T. Harper, P.D. Ryan, Palaeontol Electron. **4**, 1, (2001)
30. M.B. Kirichenko. The carabid fauna (Coleoptera, Carabidae) of the river banks, lakes shores and marshes of the Forest and Forest-Steppe of eastern part of the Ukraine. Manuscript for degree of Ph. D. (1999)
31. A. Mazzei, T. Bonacci, T. Zetto, P. Brandmayr. Nat. sicil. **23**, 1-2, 187–199 (2010).