The structure of higher aquatic vegetation in the genetic series of floodplain reservoirs

R Babko¹, T Diachenko², Y Danko³, J Zaburko⁴ and J Szulżyk-Cieplak⁵

¹ National Academy of Sciences of Ukraine, Schmalhausen Institute of Zoology, Department Fauna and Systematics of Invertebrates, B. Khmelnytsky 15, 01030 Kyiv, Ukraine
² National Academy of Sciences of Ukraine, Department of Ichthyology and Hydrobiology of River Systems, Institute of Hydrobiology, Heroiv Stalingrada 12, 04210 Kyiv, Ukraine
³ Sumy Makarenko State Pedagogical University, Department of General Biology and Ecology, Romenska 87, 40002 Sumy, Ukraine
⁴ Lublin University of Technology, Faculty of Environmental Engineering, Nadbystrzycka 40B, 20-618 Lublin, Poland
⁵ Lublin University of Technology, Faculty of Fundamentals of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland

rbabko@ukr.net, t_dyachenko@ukr.net, yaroslavdanko@gmail.com, jacek_zaburko@pollub.edu.pl, j.szulzyk-cieplak@pollub.pl

Abstract. The statistical analysis was carried out on the material collected from the Vorskla River (a first-order tributary of the Dnieper River) and the system of connected water bodies in the territory of the Getmanski National Nature Park (Sumy region, Ukraine). The methods of cluster analysis has enabled us to, first, isolate groups analogous to those obtained with the use of the method of dominants and, secondly, to compare the degree of their coincidence. From the comparison of the associations with the groups of the aquatic vegetation's species, a minimal overlapping could be seen in the composition of associations and clusters. In contrast, the groups identified by the hierarchical clustering method and by the k-means method showed significant similarities in the composition. For the objective assessment and classification of plant associations, it is most productive to use several methods with the coincidence of the results of which, it is possible to assert with more confidence that these combinations of species in nature are in fact not random.

1. Introduction

When studying vegetation, special attention is paid to the distribution of plant associations (i.e. the joint growth of groups of species) and individual plant populations in undisturbed natural sites [1]. Information on possible combinations of species in natural ecosystems is extremely important, since today, as a result of human activity, the territories of natural plant communities on all continents have been reduced to a critical minimum [2, 3, 4]. We are on the verge when some of the plants disappear, for the reasons including the lack of their natural habitats. As recent studies have shown, the rate of ongoing extinction is up to 500 times the background extinction rate for plants [5].
The problem of the conservation of aquatic plants is no less urgent, especially taking into account the extent of violation of various water bodies as a result of anthropogenic influence: changes in the riverbed morphology, pollution and pressure on hydroecosystems by seizing the production of industrial species of plants and animals [6]. Unlike terrestrial, aquatic and near-water vegetation is spatially limited because it requires a water body and has a narrow range of tolerance to the level of soil moisture. The diversity of the aquatic vegetation of natural reservoirs was largely determined by the variety of conditions that formed in the valleys of meandering rivers. It is the reclamation and management of river ecosystems that caused a reduction in the diversity of higher aquatic and near-water vegetation [7]. The issue of river renaturation is already being addressed in many European countries (e.g. England, Germany), and in the near future, is expected to dominate the scientific discourse, particularly given that the shortage of freshwater is already our present [8, 9]. Moreover, the surface water of fresh natural reservoirs is in most cases of very low quality, which is a consequence of both severe pollution of water bodies and violation of the river morphology and evolutionary established structure of their ecosystems [10].

Under these conditions, the role of aquatic vegetation is extremely high. Rivers are known to be most effectively cleaned in areas overgrown with aquatic vegetation. Plants, in addition to the ability to extract a number of biogens by accumulating them in their own biomass, are edificators of communities of numerous animal organisms. The composition of the population of thickets of aquatic vegetation is represented mainly by invertebrates and protozoa, which purify water with extreme efficiency [11]. It is on the thickets of aquatic vegetation that a trophic pyramid is based, and it is there that the passing water is most effectively conditioned.

All the elements presented in this introduction contribute to the fact that researching and studying natural associations of plants and the factors affecting their distribution becomes extremely relevant and of great practical importance.

In addition, the knowledge of the structural organisation of the plant component of aquatic ecosystems is extremely important for most European countries implementing river system restoration programmes.

2. Material and methods

2.1. Statistical analysis

Data were processed using R version 3.6.1 [12]. Hierarchical clustering was performed with hclust (function w ard.D2) from core R package stats. Two types of p-values (AU and BP) for the nodes of hierarchical clustering dendrograms were calculated with R package pvclust. According to the package documentation, “AU p-value, which is computed by multiscale bootstrap resampling, is a better approximation to unbiased p-value than BP value computed by means of normal bootstrap resampling. Clusters with AU larger than 95% are highlighted by rectangles, which are strongly supported by data” [13]. K-mean clustering was performed with function kmans from package cluster [14] and plotted on the plain of principal components with fviz_cluster from factoextra [15], non-metric multidimensional scaling (NMDS) with metaMDS from vegan [16]. Plots were produced using packages ggplot2 [17], factoextra [15], ggrepel [18], directlabels [19]. The R software environment for statistical computing and graphics is widely applied in hydrobiology and environmental engineering [20, 21, 22].

2.2. Region of study and collected data

The statistical analysis was carried out on the material collected from the Vorskla River (a first-order tributary of the Dnieper River) and the system of its associated bodies of water in the territory of the Getmanski National Nature Park (Sumy region, Ukraine). The material was collected by methods generally accepted in biocenology. In each of the water bodies, the species composition of aquatic plants of all ecological groups was determined: semi-submerged, with floating leaves, and submerged higher water plants. A percentage cover of water surface by each species of the studied water body was estimated. The obtained percentages were evaluated on a 5-point scale. Associations of aquatic plants were classified on the basis of the ecological-phytocenotic approach [23]. Data on the species
composition of the projective surface cover for 46 species of aquatic plants were obtained from 8 floodplain water bodies of various types and 1 river.

3. Results

One of the aspects of geobotanical research is the study of the spatial distribution of species, i.e. their choice of places for growth and, ultimately, the existence of any repeating cohabiting groups. Such groups are actively studied and classified and their principles of classification are different, for example, Shennikov's ecological-phytocenotic approach [23] or the popular Brown-Blanquet method [24, 25]. As for Braun-Blanquet, he restricted the use of relevés (a relevé is a list of species observed in a square plot together with estimates of their abundance/dominance or cover) to ‘homogeneous’ and ‘typical’ stands of communities. As a result, the relevé method is burdened by much bias, subjectivity, inconsistency, arbitrariness, circular argumentation and large sampling error [26, 27, 28]. Possibly the problems associated with sampling can be resolved by the use of various objective sampling designs and standardised square plot sizes [28, 29]. Moreover, the classification based on co-growth is not successful for all types of ecosystems. For terrestrial ecosystems, the most objective groupings are for the ecosystems dominated by woody vegetation (forests). The distribution of species of aquatic vegetation is determined by factors other than those that are important for terrestrial vegetation. The distribution of aquatic and near-water vegetation (hygrophilic) most depends on such factors as the speed of the current, the intensity of water exchange in weakly flowing water bodies, and the quality of the aquatic environment. A significant role is played by the content of nutrients and trace elements.

The classification of plant associations is according to the dominant species of each vegetation tier. Such associations do not account for the entire ecosystem but subjectively, based on the study of local sites. Thus, the allocation of associations without the use of statistical methods is largely subjective, which causes that its results debatable. This method involves the identification of combinations of species and is not aimed at typifying water bodies. The distinguished associations are objective, as their existence is results from specific conditions determined by the type of reservoir: whether it is a flowing body of water or not and as well as what trophic status the water body has. In other words, there can be different associations in lakes and rivers, and water bodies can be systematised on their basis. The adaptation of individual species to the water bodies of different types can potentially serve as the basis for their classification. Using factual material regarding the distribution and quantitative representation of species of hygrophilic plants in reservoirs of various types, we systematised reservoirs by the clustering method (Figure 1).

**Figure 1.** Ward's hierarchical clustering of objects based on Kulczynsky distances matrix of Hellinger-transformed data. AU (Approximately Unbiased) and BP (Bootstrap Probability) p-values are shown next to the nodes. For the codes of locations see Table 1.
According to the results obtained, most of the water bodies with high reliability combined with the river, which shows that the river and a number of its genetic derivatives form a kind of unity. At the same time, two lakes stood out in a separate cluster. Both lakes are specific because despite the shallow depth — from 1 to 3 m — the water column in them is stratified and the lower layers of the water and bottom sediments are anaerobic. Conventionally, we designate this category of reservoirs in the river valley as “hydrogen sulphide lakes”. This type of reservoir does not grow idiosyncratic plant species and is distinguished by their small number.

**Figure 2.** Ward's hierarchical clustering of species based on a Kulczynski distances matrix of Hellinger-transformed data. AU (Approximately Unbiased) and BP (Bootstrap Probability) p-values are shown next to the nodes. Abbreviations of plant species: Acorcala — Acorus calamus, Aligram — Alisma gramineum, Alisaqua — Alisma plantago aquatica, Batrfoen — Batrachium foeniculaceum, Bidefron — Bidens frondosa, Bidetrip — Bidens tripartita, Butoumbe — Butomus umbellatus, Cerademe — Ceratophyllum demersum, Cerasubm — Ceratophyllum submersum, Elodcana — Elodea canadensis, Glycermaxi — Glyceria maxima, Hipplanc — Hippuris lanceolata, Hottpalu — Hottonia palustris, Hydranna — Hydrocharis morsus ranae, Irispseu — Iris pseudacorus, Lemnmino — Lemna minor, Lemntris — Lemna trisulca, Mentqua — Mentha aquatica, Myospalu — Myosotis palustris, Myrispic — Myriophyllum spicatum, Myrivert — Myriophyllum verticillatum, Nuphlute — Nuphar lutea, Nympcand — Nymphaea candida, Oenaqua — Oenanthe aquatica, Phraaust — Phragmites australis, Poliamph — Polygonum amphibium, Potacomp — Potamogeton compressus, Potacrys — Potamogeton crypsus, Potaluce — Potamogeton lucens, Ptmgtnmt — Potamogeton natans, Potapect — Potamogeton pectinatus, Potaperf — Potamogeton perfoliatus, Ranulinq — Ranunculus lingua, Rumehydr — Rumex hydrolapatum, Sagisagi — Sagittaria sagittifolia, Scirlacu — Scirpus lacustris, Siumlati — Sium latifolium, Sparemer — Sparganium emersum, Sparerec — Sparganium erectum, Spirpoly — Spirodela polyrrhiza, Straaloi — Stratiotes aloides, Sympoffi — Symphytum officinale, Typhangu — Typha angustifolia, Typhlati — Typha latifolia, Utricularia minor, Utrivulg — Utricularia vulgaris.
The hierarchical clustering of species based on the data on the distribution of all identified species in the studied reservoirs showed that their entire array is divided with a high level of reliability into 10 clusters, each representing 20 species (Figure 2). An analysis of the composition of this cluster showed that these are mainly species found in flowing waters.

The comparison of the associations established by the method of dominants with the groups of the aquatic vegetation's species, obtained from cluster analysis, showed a minimal overlapping in the composition of associations and clusters.

Furthermore, non-hierarchical k-means clustering was applied. This method requires presetting the desired number of clusters. To determine the optimal number of clusters, we used two methods: silhouette and Calinski-Harabasz. According to the silhouette method, the optimal number of clusters is 7, while according to Calinski-Harabasz — 2. A number of clusters 7 in our case appears, then, redundant. To visualise the clusters, they were depicted on the plane of the first two principal components of the PCA. The k-means clustering with 2 preset clusters showed that the floodplain PMK lake joined to the hydrogen sulphide lakes (Figure 3). Considering that the PMK Lake is stratified, most of the water column is oxygen-free, and the bottom sediments are represented by sapropel, whose presence in a reservoir in this group seems perfectly logical.

Both clustering methods show that in the genetic series of floodplain bodies of water, lakes in which secondary pollution causes eutrophication and triggers bogging processes in quantitative representation and number of species differ from reservoirs that maintain a connection with the river.

In the case of species, the silhouette method recommends for k-means clustering to choose seven clusters (Figure 4B), the Calinski-Harabasz method — two clusters (Figure 4A). The blue cluster (Figure 4A) united species that are tolerant to waterlogging. For our data, the most acceptable number of clusters is seven. The results of clustering make it possible to isolate groups similar to groups established using the method of dominants and compare the degree of their coincidence. In our case, the use of the k-means method allows us to make the partition into groups more objective. The results are shown in Figure 4B. Each of the obtained seven clusters combined the species closest in environmental requirements, with similar reactions to limiting factors including the degree of water exchange.
Figure 4. K-mean clustering of species on the plain of the first two principal components. Data were A – Calinski-Harabasz transformed and B – Hellinger-transformed. For codes of species see Figure 2.

4. Discussion
The grouping of aquatic vegetation based on the dominants method was tested by two statistical methods. Incorporation of statistical methods allows us to maximally avoid subjectivity in interpreting the results of field studies. According to the dominants technique, the following associations were identified in each of the examined bodies of water (Table 1).

Table 1. The associations of aquatics plants from Vorskla River and the floodplain lakes according to dominants classification.

| Type of water body (Code) | Associations and their numbers |
|--------------------------|--------------------------------|
| Vorskla river (Rv_Vrskl) | Sagittaria sagittifolia        |
|                          | Nuphar lutea                   |
|                          | Potamogeton perfoliatus        |
|                          | Ceratophyllum demersum         |
| Backwater Zhuravne (BW_Zh) | 1 Ceratophyllum demersum      |
|                          | Spirodes polyrrhyza            |
|                          | Nuphar lutea                   |
|                          | Ceratophyllum demersum         |
|                          | Lemna minor                    |
|                          | Sagittaria sagittifolia        |
|                          | Nymphaea candida               |
|                          | Lemna minor                    |
| Backwater Monastyrskie (BW_M) | 1 Ceratophyllum demersum      |
|                          | Spirodes polyrrhyza            |
|                          | Nuphar lutea                   |
|                          | Ceratophyllum demersum         |
|                          | Sagittaria sagittifolia        |
|                          | Nymphaea candida               |
| Backwater Bujmerovka (BW_Bu) | 1 Elodea canadensis           |
|                          | Lemna trisulca                 |
|                          | Lemna minor                    |
|                          | Nuphar lutea                   |
|                          | Ceratophyllum demersum         |
|                          | Lemna trisulca                 |
|                          | Sagittaria sagittifolia        |
|                          | Ceratophyllum demersum         |
|                          | Elodea canadensis              |
|                          | Ceratophyllum demersum         |
| Lake Bujmerovka-1 (Lk_Bm1) | Nuphar lutea                   |
|                          | Stratiotes aloides             |
|                          | Lemna minor                    |
|                          | Ceratophyllum demersum         |
|                          | Spirodes polyrrhyza            |
|                          | Lemna trisulca                 |
| Lake Bujmerovka-2 (Lk_Bm2) | 1 Ceratophyllum demersum      |
|                          | Lemna minor                    |
|                          | Lemna trisulca                 |
| Lake Monastyrskie (Lk_Mon) | 1 Ceratophyllum demersum      |
|                          | Spirodes polyrrhyza            |
|                          | Lemna minor                    |
|                          | Stratiotes aloides             |
|                          | Lemna minor                    |
| Lake PMK (Lk_PMK) | 1 Ceratophyllum demersum      |
|                          | Spirodes polyrrhyza            |
|                          | Nuphar lutea                   |
|                          | Nymphaea candida               |
|                          | Typha angustifolia             |
|                          | Ceratophyllum demersum         |
| Lake H2S (Lk_H2S) | 1 Lemna trisulca               |
|                          | Lemna minor                    |
|                          | Lemna minor                    |
|                          | Lemna trisulca                 |
|                          | Typha angustifola              |
|                          | Phragmites australis           |
|                          | Typha angustifolia             |
|                          | Scirpus lacustris              |
|                          | Lemna trisulca                 |
|                          | Lemna minor                    |
|                          | Lemna trisulca                 |


An analysis of the groups identified by the hierarchical clustering method and the k-means method showed significant similarities in the composition of these groups. Therefore, the composition of two groups coincided by 100% (Table 2). The first and second clusters obtained by hierarchical clustering coincided with the groups identified by the k-means method. Clusters 3 and 4 also formed a common group at a higher level of unification with group 3, isolated by the k-means method.

The coincidence of the groups identified by three different methods was observed in only one case. This association is represented by species from one ecological group — plants with floating leaves and represented by species *Nymphaea candida* and *Nuphar lutea*. It should be noted that the k-means method included two more species in the composition of this group: *Spirodea polyrrhyza* and *Phragmites australis*. A cluster obtained by hierarchical clustering fully coincided with the dominants association. Given that the dominants method allowed 23 associations to be distinguished, its coincidence with the clustering method amounted to approximately 5%.

### Table 2. Plant species grouped of the basis of different statistical techniques.

| Hierarchical clustering | K-mean | Braun-Blanket |
|-------------------------|--------|---------------|
| Cluster | Name species | Group | Name species | Association | Name species |
| 1 | *Lemna trisulca* | 1 | *Lemna trisulca* | *Lemna minor* | *Lemna minor* |
| | *Lemna minor* | 1 | *Lemna minor* | *Hydrocharis morsus ranae* | *Hydrocharis morsus ranae* |
| | *Hydrocharis morsus ranae* | 1 | *Hydrocharis morsus ranae* | *Ceratophyllum demersum* | *Ceratophyllum demersum* |
| 2 | *Typha angustifolia* | 2 | *Typha angustifolia* | *Spirodela polyrrhyza* | *Spirodela polyrrhyza* |
| | *Spirodela polyrrhyza* | 2 | *Spirodela polyrrhyza* | *Sagittaria sagittifolia* | *Sagittaria sagittifolia* |
| 3 | 3 | *Sagittaria sagittifolia* | *Potamogeton natans* | *Potamogeton natans* | *Myriophyllum spicatum* | *Myriophyllum spicatum* |
| 4 | *Sparganium emersum* | 4 | *Sparganium emersum* | *Butomus umbellatus* | *Butomus umbellatus* |
| 5 | *Nymphaea candida* | 4 | *Nymphaea candida* | *Nymphaea candida* | *Nymphaea candida* |
| | *Nuphar lutea* | 1 | *Nuphar lutea* | *Nuphar lutea* | *Nuphar lutea* |

### 5. Conclusion

Despite the widespread use of the dominants method for classification of aquatic plants communities, it should be noted that its results frequently do not coincide with the results of even simple statistical analysis. Apparently, the allocation of groups by the dominants method is associated with a high level of subjectivity.
It is likely that for an objective assessment and classification of plant associations, it is most productive to use several methods the combined results of which allow for a more confident assertion that these combinations of species in nature are, in fact, not random.

6. References

[1] Craft C, Krull K and Graham S 2007 Ecological indicators of nutrient enrichment, freshwater wetlands, Midwestern United States (US) *Ecol. Indic.* 7 (4) 733–50

[2] Camargo J A, Alonso A 2006 Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment *Environ. Int.* 32 831–49.

[3] Huijbregts M A J 2011 Implementing groundwater extraction in life cycle impact assessment: Characterization factors based on plant species richness for the Netherlands *Environ. Sci. Technol.* 45 (2) 629–35

[4] Lambert S J, Davy A J 2011 Water quality as a threat to aquatic plants: discriminating between the effects of nitrate, phosphate, boron and heavy metals on charophytes *New Phytol.* 189 1051–59

[5] Humphreys A M, Govaerts R, Ficinski S Z, Lughadha N E and Vorontsova M S 2011 Implementing groundwater extraction in life cycle impact assessment: Characterization factors based on plant species richness for the Netherlands *Environ. Sci. Technol.* 45 (2) 629–35

[6] Lamberti S J, Davy A J 2011 Water quality as a threat to aquatic plants: discriminating between the effects of nitrate, phosphate, boron and heavy metals on charophytes *New Phytol.* 189 1051–59

[7] Humphreys A M, Govaerts R, Ficinski S Z, Lughadha N E and Vorontsova M S 2011 Implementing groundwater extraction in life cycle impact assessment: Characterization factors based on plant species richness for the Netherlands *Environ. Sci. Technol.* 45 (2) 629–35

[8] Lamberti S J, Davy A J 2011 Water quality as a threat to aquatic plants: discriminating between the effects of nitrate, phosphate, boron and heavy metals on charophytes *New Phytol.* 189 1051–59

[9] Kaenel B R, Buehrer H and Uehlinger U 2000 Effects of aquatic plant management on stream metabolism and oxygen balance in streams *Freshw. Biol.* 45 (1) 85–95

[10] Kim J H, Chung H Y, Kim S H, Kim J G 2016 The influence of water characteristics on the aquatic insect and plant assemblage in small irrigation ponds in Civilian Control Zone *Korea J. Wet. Res.* 18(4) 331–41

[11] Milà I Canals L, Chenoweth J, Chapagain A, Orr S, Antón A and Clift R 2009 Assessing freshwater use impacts in LCA: Part I Inventory modelling and characterisation factors for the main impact pathways int *J. Life Cycle Assess* 14(1) 28–42

[12] Pfister S, Koehler A and Hellweg S 2009 Assessing the environmental impacts of freshwater consumption in LCA *Environ. Sci. Technol.* 43(11) 4098–104

[13] R Core Team 2019 *R*: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna: Austria) https://www.R-project.org/.

[14] Maechler M, Rousseeuw P, Struyf A, Hubert M and Hornik K 2019 *Cluster*: Cluster Analysis Basics and Extensions. R package version 2.0.8.

[15] Kassambara A, Mundt F 2017 *factoextra*: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. https://CRAN.R-project.org/package=factoextra

[16] Wickham H 2009 *ggplot2*: elegant graphics for data analysis. Springer New York.

[17] Slowikowski K 2018 *ggrepel*: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.0. https://CRAN.R-project.org/package=ggrepel

[18] Hocking T D 2018 *directlabels*: Direct Labels for Multicolor Plots. R package version 2018.05.22. https://CRAN.R-project.org/package=directlabels
[20] Pliashechnyk V, Danko Y, Łagód G, Drewnowski J, Kuzmina T and Babko R 2018 Ciliated protozoa in the impact zone of the Uzhgorod treatment plant E3S Web of Conferences 30 02008. DOI: 10.1051/e3sconf/20183002008

[21] Babko R, Szulżyk-Cieplak J, Danko Y, Duda S, Kirichenko-Babko M and Łagód G 2019 Effect of stormwater system on the receiver Journal of Ecological Engineering 20(6) 52–59 https://doi.org/10.12911/22998993/109433

[22] Łagód G, Duda S M, Majerek D, Szutt A and Dolhańczuk-Śródka A 2019 Application of electronic nose for evaluation of wastewater treatment process effects at full-scale WWTP Processes 7 251 doi:10.3390/pr7050251

[23] Shennikov A P 1964 Vvedenie v geobotaniku. Leningrad LGU [in Russian]

[24] Westhoff V and Maarel E van der. 1973 The Braun-Blanquet approach: Handbook of vegetation science Hague 5, 619–726

[25] Neuhauslova Z et al. 2001 Potential natural Vegetation of the Czech Republic. II Braun-Blanquetia, Camerino 30 80

[26] Chytrý M 2001 Phytosociological data give biased estimates of species richness J. Veg. Sci. 12 439–44

[27] Chytrý M and Otypková Z 2003 Plot sizes used for phytosociological sampling of European vegetation J. Veg. Sci. 14 563–70

[28] Podani J 2006 Braun-Blanquet’s legacy and data analysis in vegetation science J. Veg. Sci. 17 113–17

[29] Bruelheide H and Chytrý M 2000 Towards unification of national vegetation classifications: A comparison of two methods for analysis of large data sets J. Veg. Sci. 11 295–306.