Multidimensional Statistical Analysis of the Structure of Planktonic and Bottom Communities Inhabiting Mineralized Rivers of the Elton Lake Basin

T. D. Zinchenko[a], *, V. K. Shitikov[a], L. V. Golovatyuk[a], and E. V. Abrosimova[a]

[a] *Institute of Ecology of the Volga River Basin, Russian Academy of Sciences, Togliatti, 445003 Russia*
*e-mail: zinchenko.tdz@yandex.ru

Abstract—This paper presents results of comprehensive studies of saline rivers in the arid Elton Lake region based on hydrobiological surveys of planktonic and bottom communities. Two tables representing the species structure of communities and produced on the basis of survey data collected on 13 river sites in 2013 and 2018, including data on the abundance of hydrobionts belonging to 94 different taxonomic groups of macrozoobenthos, meiobenthos, and zooplankton, have been compared. The co-inertia analysis has identified a high statistical consistency of the data matrices determined by objective spatial distribution patterns of aquatic organisms. A randomization test was applied to the Procrustean correlation coefficient, and it showed that the conjugacy of the two species structures in the space of latent variables is statistically significant ($p = 0.00026$). Concurrently, a change trend was noted for the taxonomic composition of some surveyed communities affected by dynamic abiotic factors. The relationship between the taxonomic structure of hydrobionts and the combination of 30 abiotic environmental factors measured in the course of the hydrobiological and hydrochemical monitoring of the studied river sites was analyzed. Using canonical correlation analysis and the technique of projection to latent structures, a complex of ordination diagrams has been produced permitting the identification of specific "environmental niches" for each species group, including certain sets of biotope characteristics. A significant mutual correlation between the planktonic and bottom communities has been established. This correlation indicates close interrelations between these communities determined both by biotic interactions and by mutually consistent reactions to changes in aquatic environmental factors.

Keywords: saline rivers, zooplankton, macrozoobenthos, meiobenthos, community structure, co-inertia analysis, Procrustean analysis, multidimensional ordination

DOI: 10.1134/S1062359020100283

INTRODUCTION

The region of Elton Lake is a unique natural and territorial complex of the Lower Volga River basin. The development of meso- and polyhaline river systems within its boundaries is affected by stochastic and, sometimes, extreme natural and climatic disturbances typical for arid zones amid concurrent impacts of anthropogenic factors (Zinchenko et al., 2017). Such disturbances cause both sharp fluctuations in the abundance of individual populations and regular changes in the complex of predominant species in hydrobiont communities.

The statistical analysis of the taxonomic structure of planktonic and bottom communities inhabiting saline rivers in the Elton Lake basin (Volgograd region) earlier performed by the authors using various multidimensional methods made it possible to identify specific consortium-type associations that are dynamically forming according to the adaptive capabilities of

individual species (Zinchenko et al., 2018). It was shown that life forms of the planktonic and bottom communities are not clearly distinguished in the ecosystems studied; i.e., the transition of benthic animals to the plankton lifestyle was noted, which results in a high share of interpenetrating environmental groups. The strength of such associative intrasystem relationships was analyzed using a special algorithm integrating multidimensional ordinations of macrozoobenthic, meiobenthic, and zooplanktonic communities (Shitikov and Zinchenko, 2019). The assessment of the canonical correlation between the individual data blocks indicates a close relationship between these groups determined both by biotic interactions and by mutually consistent reactions to changes in aquatic environmental factors.

The above studies were performed on the basis of a complex hydrobiological survey carried out in August 2013. After the preparation of a similar block of data on

Table 1. Rivers of the Elton Lake basin, location of stations, and key hydrochemical parameters

Rivers	Stations	River site	Hardness, mg-equiv/L	Total mineralization, g/L	Chlorides (Cl^-), g/L	Oxygen (O_2), mg/L	Phosphate ions (PO_4^-), mg/L	Ammonium (NH_4^+), mg/L
Solyanka	Sol_1	Upper	105.0	27.6	11.4	6.5	0.6	5.1
	Sol_3	Estuary	109.0	27.7	16.9	7.2	0.5	3.5
Lantsug	Lan_1	Middle	24.0	50.3	28.4	7.1	2.1	0.6
	Lan_2	Estuary	78.0	13.4	6.6	7.5	1.4	3.6
Khara	Khar_2	Upper	122.0	20.0	11.4	6.7	0.7	0.4
	Khar_3	Middle	99.5	17.7	8.7	7.0	0.7	2.1
	Khar_6	Lower	77.0	15.3	6.0	7.4	0.8	3.1
	Khar_7	Estuary	90.5	15.9	7.6	6.1	0.4	3.3
Chernavka	Cher_1	Upper	120.0	28.9	16.9	6.5	0.4	3.5
	Cher_2	Middle	118.0	28.8	16.8	6.4	0.5	3.1
	Cher_3	Estuary	121.0	30.5	18.1	7.6	0.9	3.8
Bolshaya Samoroda	B.S._2	Middle	42.4	11.1	5.3	7.9	2.2	0.3
	B.S._3	Estuary	56.0	13.2	6.2	7.7	2.2	0.3

The analytical study of the hydrochemical water samples was conducted by the certified laboratory "Center for Monitoring of Aquatic and Geological Environment," Samara.

the basis of a survey carried out in 2018, a number of actual theoretical and practical issues have been identified:

(1) How can we assess quantitatively the intensity of changes in the species structure for the two blocks of parallel hydrobiological observations carried out in different years, and how can we check the hypothesis of statistically significant differences between them?

(2) What statistical method is most suitable to model the relationship between the species structure of hydrobionts and the extended complex of environmental variables?

(3) Is it possible to reduce, without significant information losses, the initial space of variables through the selection of a limited set of species that explain the spatial heterogeneity of biocenoses to the maximum extent, and is it possible to identify the key environmental factors forming the biotope conditions?

The purpose of this study was to assess the species structure of planktonic and bottom communities affected by a complex of interrelated abiotic factors using various methods of statistical analysis.

MATERIALS AND METHODS

The source data have been collected in the course of hydrobiological surveys of the macrozoobenthos, meiobenthos, and zooplankton carried out on various sites of five saline rivers (Table 1) in the Elton Lake basin ((49°07′30″ N, 46°30′40″ E) having significant mineralization gradients (5.3 to ≥50 g/L). Detailed descriptions and map schemes of the study area, hydrobiont sampling techniques, and data processing methods used in the laboratory environment are provided in earlier works by T.D. Zinchenko et al. (Zinchenko et al., 2010, 2019; Zinchenko et al., 2017, 2018). The results of the hydrobiological surveys simultaneously carried out at 13 stations on the Lantsug, Khara, Chernavka, Bolshaya Samoroda, and Solyanka rivers in August 2018 were used in the statistical analysis. A table representing the abundance of 94 taxonomic groups, including 30 genera and species of macrozoobenthos, 33 meiobenthos (ind./m^2), and 31 zooplankton (ind./m^3), was produced.

To ensure the correct cooperative data processing, abundance values have been converted into a unified 0 to 6 scale; their optimal gradations were identified using the k nearest neighbors algorithm (Shitikov and Rozenberg, 2014). Two taxonomic abundance matrices produced on the basis of the hydrobiological surveys carried out in 2018 (HB$_{18}$, 13 × 94 in size) and in 2013 (HB$_{13}$, 13 × 88 in size (Zinchenko et al., 2018)) were used as the source data. Abbreviations of the river sites are provided in Table 1; the taxa codes have the following prefixes: ZB (macrozoobenthos), MB (meiobenthos), and ZP (zooplankton) (Table 2).

Table 2. Groups of hydrobiont taxa in rivers of the Elton Lake basin distinguished on the basis of characteristic habitat features and key abiotic factors strongly correlating with these groups

Group	Taxon	Code	T	R^2	Abiotic factors
I	*Brachionus quadridentatus*	ZPBra_qd	2	0.704	$NA^+ + Ka^+$ HCO_3^- Depth, h Cl^-
	Cricotopus rufiventris	MBCri_rf	2	0.614	
	Microchironomus deribae	ZBMic_tn	2	0.515	
	Brachionus calyciflorus	ZPBra_cl	1	0.710	
	Chironomus plumosus	ZBChi_pl	1	0.710	
II	*Colurella adriatica*	ZPCol_ad	4	0.916	$PO_4^{3-} - PO_2$
	Acanthocyclops americanus	ZPAca_am	2	0.773	
	Cricotopus rufiventris	ZBCri_rf	2	0.560	
	Gammarus lacustris	ZBGam_lc	3	0.565	
	Ethmolaimus multipapillatus	MBEth_ml	2	0.598	
	Candoninae gen. spp.	MBCan_sp	1	0.579	
	Chironomus aprilinus	ZBChi_ap	4	0.422	
III	*Glyptotendipes salinus*	ZBGly_sl	3	0.568	CO_3^- pH Cu
	Testudinella patina	ZPTes_pt	2	0.795	
	Odontomyia sp.	ZBOdn_sp	3	0.675	
	Tanytarsus kharaensis	ZPTan_kh	2	0.821	
	Cyclopoida gen. spp. *nauplii*	ZPNau_cl	12	0.172	
	Megacyclops viridis	ZPMeg_vr	4	0.399	
	Cletocamptus confluens	ZPCle_cn	2	0.852	
	Hygrotus enneagrammus	ZBHyg_en	2	0.508	
	Cricotopus sylvestris	ZBCri_sv	2	0.642	
	Cletocamptus confluens	MBCle_cn	1	0.711	
	Tanytarsus kharaensis	ZBTan_kh	1	0.711	
	Heterocypris salina	MBHet_sl	2	0.734	
IV	*Chironomus salinarius*	ZBChi_sl	6	0.567	Total hardness Ca^{2+} Mg^{2+} BOD_{20}
	Monhystrella parvella	MBMon_pv	12	0.493	
	Sigara sp.	ZBSig_sp	3	0.525	
	Cletocamptus retrogressus	MBCle_rt	8	0.540	
	Brachionus plicatilis	ZPBra_pl	13	0.539	
	Cricotopus salinophilus	ZBCri_sf	9	0.508	
	Cletocamptus retrogressus	ZPCle_rt	10	0.588	
V	*Apocyclops dengizicus*	ZPApo_dn	4	0.549	NH_4-N Zn CI"a"
	Cricotopus salinophilus	MBCri_sf	6	0.622	
	Apocyclops dengizicus	MBApo_dn	5	0.635	
	Cricotopus salinophilus	ZPCri_sf	10	0.712	
	Palpomyia schmidti	ZBPal_sh	5	0.749	
	Cyprideis torosa	MBCyp_tr	8	0.287	

(T) occurrence on the river sites, (R^2) length of the correlation vector in Fig. 2c; codes of the registered taxa have the following prefixes: ZB (macrozoobenthos), MB (meiobenthos), and ZP (zooplankton).
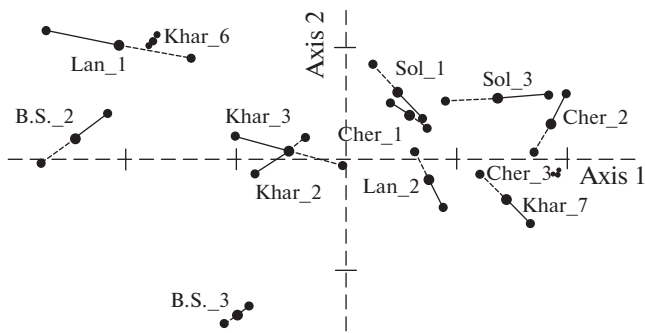
Fig. 1. Drifts of the projections of river sites according to hydrobiological surveys carried out in 2013 and 2018 relative to the consensus configuration in the coordinates of the two latent structure axes based on the Procrustean co-inertia analysis. See Table 1 for the key to river sites.

Hydrochemical monitoring was concurrently carried out at the same stations; the HC_{18} matrix 13×30 in size of the monitored environmental factors, including hydrological parameters of water streams (depth, flow velocity, etc.) and hydrophysical and hydrochemical parameters of waters in the five saline rivers, has been produced on its basis.

The symmetrical co-inertia analysis (CIA) (Dolédec and Chessel, 1994) was used to compare the blocks of data obtained in the course of the simultaneous hydrobiological surveys carried out in 2013 and 2018; this analysis is considered the most general conjugacy assessment method for two matrices describing taxonomic structures of communities or various ordinations produced on their basis. The latent variables formed are directed towards the maximum multidimensional inertia (an information measure defined as the statistical value of the Pearson's chi-squared test (χ^2) for a conjugacy table produced for the two associated tables).

The CIA-based Procrustean co-inertia analysis was performed to compare dot configurations of the two parallel diagrams using the 'superimposition' method (Dray et al., 2003). For that purpose, the space geometry of the two ordinations was brought to a consistent shape using scaling, centering, and rotation. Ultimately, the Procrustean correlation coefficient was calculated: $R_p = \sqrt{1 - m^2}$, where m^2 is the sum of squared distances between the coordinates of dots on the two ordinations.

The model describing the relationship between the taxonomic structure of hydrobionts (Y) and a combination of abiotic factors (X) was built on the basis of the partial least squares (PLS) method (Wold, 1985), which is a generalization of the classic least squares method with respect to canonical correlations. At the first stage, the PLS projects the initial matrices to the new reduced space and then distinguishes a set of latent variables that explain the covariance maximum between the response (Y) and predictors (X). The assessment vector for the B coefficients in the regression $Y = XB$ is computed on the basis of orthogonal estimates and loads on the axes of the latent structure formed. With regards to the multidimensional ordination, the PLS regression predicts the values of the main coordinates of the objects analyzed (i.e., river sites).

The modernized principal components analysis (srPCA (Lê Cao et al., 2008)) was used to distinguish latent variables from the initial matrices. The inclusion of an additional regularization member makes it possible to solve the issue of ill-conditioned covariance matrices, while the sparse algorithm components perform the selection of the informative complex of variables most closely related to the response.

The transposition test was used to assess the statistical significance of the Procrustean correlation, while the adequacy of the models produced was verified using the cross-check algorithm (Shitikov and Rozenberg, 2014).

The statistical analysis was performed using the ade4 and mixOmics software packages (González et al., 2012; Singh et al., 2016) in the R programming environment, version 3.06. Additional information on the methods used in the study can be found in a number of overviews (Shitikov and Zinchenko, 2019; Hervé, 2016; Hervé et al., 2018) and in the blog of one of the authors (https://stok1946.blogspot.com/2019/03/blog-post.html).

RESULTS AND DISCUSSION

The co-inertia analysis (CIA) was used to assess the spatial distribution dynamics of the species composition in the period of 2013−2018 on river sites with different mineralization levels and other biotope characteristics. For this purpose, a consensus ordination diagram of the stations has been built on the basis of the generalized species space (88 + 94) produced as a result of the merger of the HB_{13} and HB_{18} hydrobiological survey matrices. The two latent structure (LS) axes built explained 59.6% by the combined statistical variation and identified a high similarity between the compared ordinations. Figure 1 presents a Procrustean analysis diagram of drifts of the coordinates of individual river sites: the initial positions of biotopes (based on the survey carried out in 2013) are shown at one end of each segment; their positions in 2018 are shown at the opposite end of each segment (connected by the dashed line), while the consensus (i.e., averaged) states of the species structures are shown by heavy dots in the middle. The correlation coefficients

between the consensus ordination axes and latent variables are 0.969 and 0.971, respectively; while the *RV* coefficient representing the generalized similarity measure for the two data tables is equal to 0.838.

The statistical significance of the interdetermination of the two observation data blocks was assessed on the basis of the Procrustean correlation coefficient (R_p) using a randomization procedure. In the course of this procedure, the table rows were randomly mixed up 999 times, and the null distribution of the R_p value was produced assuming that the H_0 hypotheses is true. Because the empirical value ($R_p = 0.94$) significantly exceeds the average value ($R_p = 0.87$) of the imitated distribution, the null hypothesis stating that the relationship between the HB_{13} and HB_{18} tables is coincidental was rejected at the significance level: $p = 0.00026$.

It is necessary to note that amid the preservation of the main patterns of the species spatial distribution, the taxonomic composition at individual stations changed over time to a greater or lesser degree; i.e., dynamic combinations of interacting taxa were observed. For instance, in 2013, the complex of predominant taxa dependent on the spatial location of biotopes to the maximum extent included *Tanytarsus kharaensis* (ZB), *Microchironomus deribae* (ZB), *Nais elinguis* (ZB), *Nitokra lacustris* (MB), *Cletocamptus confluens* (MB), *Cletocamptus confluens* (ZP), *Lecane hamata* (ZP), and *Cyprideis torosa* (ZP); while in 2018, it included *Tanytarsus kharaensis* (ZB), *Cricotopus salinophilus* (ZB), *Palpomiya schmidti* (ZB), *Chironomus salinarius* (ZB), *Chironomus aprilinus* (ZB), *Cricotopus rufiventris* (ZB), *Cletocamptus confluens* (MB), *Cletocamptus confluens* (ZP), *Alona rectangula* (ZP), and *Cricotopus salinophilus* (ZP).

The relationship between the taxonomic structure of hydrobionts represented by HB_{18} and the combination of abiotic environmental factors represented by the HC_{18} matrix was identified through the production of ordination models on the basis of the PLS regression. The models were produced in two formats: using the entire complex of initial variables and selecting the set of the most influencing predictors. The potential loss of information due to the reduced space of attributes was assessed according to the methodology described above, i.e., by assessing the statistical significance of the Procrustean correlation coefficient (R_p) between the tables prior to and after the reduction. It turned out that the reduction in the number of environmental parameters contained in the HC_{18} matrix from 30 to 21 and the reduction in the taxa number from 94 to 40 result in negligibly small changes in the configuration of the resultant diagrams.

Figures 2a−2c show three interrelated graphs reflecting the reduced ordination model. The 2a diagram illustrates the ordination of the 13 river sites determined both by the hydrobiont species composition in each biocenosis and the combination of hydrochemical conditions in each biotope. The sites on the graph are grouped by their general mineralization gradations: (1) over 25 g/L, (2) 15−25 g/L, and (3) less than 15 g/L. It is possible to assume that the first ordination axis relates to the water mineralization factor, while the second axis accumulates the effects of the other environmental variables.

Figure 2b shows the correlation relationship between the abiotic parameters and the axes of the ordination produced in more detail. The results presented indicate that the coordinates of the river sites shown in Fig. 2a (and, accordingly, the species composition in each community) are determined not only by the total mineralization, but by a broad range of other environmental factors as well. For instance, the negative direction of axis 1 represents higher water hardness levels, including high concentrations of calcium and magnesium ions that determine a cluster of sites on the highly mineralized Solyanka and Chernavka rivers. By contrast, axis 2 is primarily determined by the sodium chloride and hydrocarbonate concentrations; high values of these parameters are typical for the upper reaches of the Lantsug River. Sites on the Bolshaya Samoroda River feature low concentrations of the above components but relatively high concentrations of dissolved oxygen and phosphates. It is also necessary to note that concentration fluctuations of copper ions are close to the instrumental error and the strong correlation of this parameter should likely be considered "false."

Figure 2c reflects the correlation relationships between the abundance of individual hydrobiont taxa and the axes of the latent structure. The superimposition of correlation circles 2b and 2c makes it possible to identify aquatic environmental factors creating tolerant conditions for each species included in the analysis. Based on the similarity analysis of correlation coefficients, five relatively characteristic species groups have been distinguished (I–V in Fig. 2c), and the key hydrochemical parameters that determine their biotopic requirements to the maximum extent from the multidimensional ordination perspective have been established for each of these groups. The analysis of the groups distinguished makes it possible to state that in non stationery conditions of highly mineralized rivers, planktonic and bottom communities form a jointly functioning dynamic consortium-type association whose species are united both by the same tolerance to aquatic environmental factors and common trophic relations with the substrate.
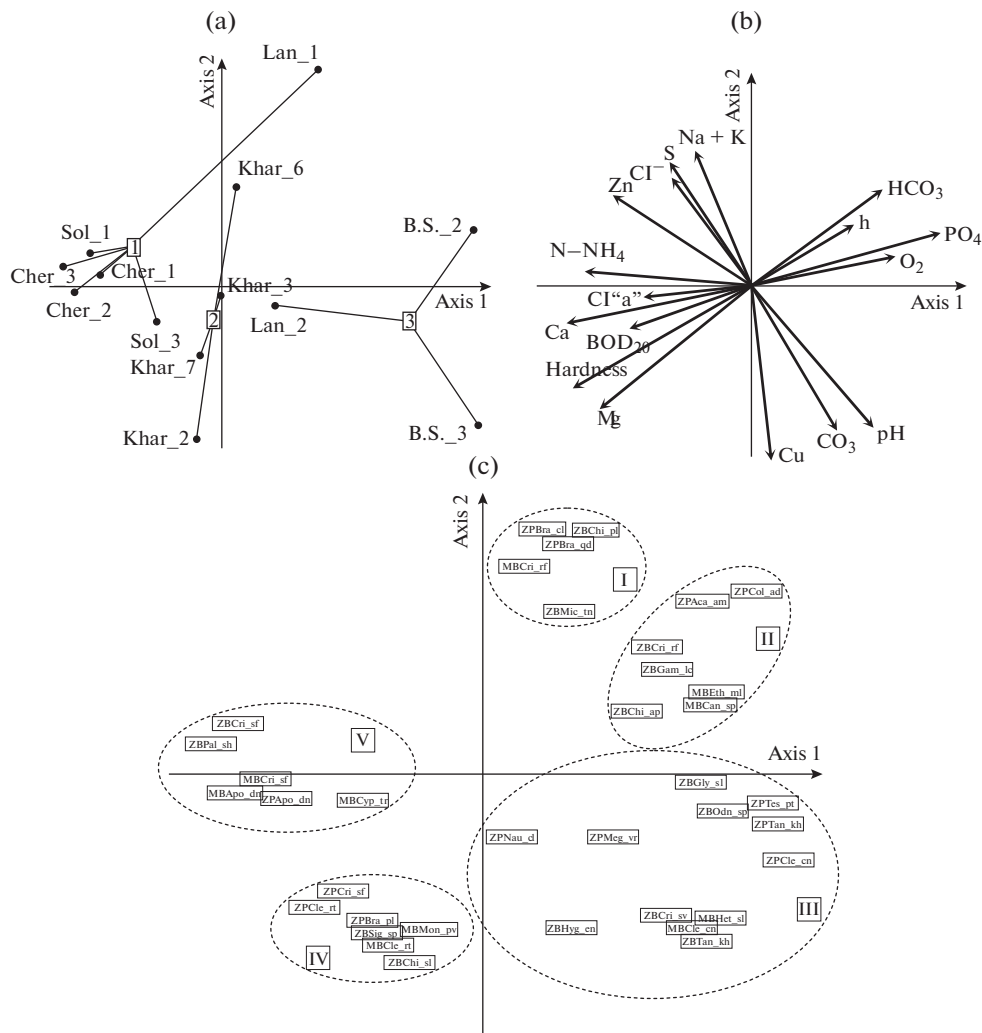
Fig. 2. Results of the direct ordination analysis of the relationship between the taxonomic structure of hydrobionts inhabiting rivers of the Elton Lake basin and abiotic environmental factors performed using the partial least squares method: (a) ordination of river sites in the space of two latent structure (LS) axes (see Table 1 for the key to the river sites); (b) diagram showing the correlation between the abiotic factors and LS axes; (c) diagram showing the correlation between the abundance of individual aquatic taxa and LS axes (see Table 2 for the taxa codes).

CONCLUSIONS

Planktonic and benthic communities inhabiting saline rivers of the Elton Lake basin demonstrate sustainable patterns in the spatial distribution of species abundance values. A comparison of the two matrices representing the results of hydrobiological surveys carried out in 2013 and 2018 demonstrates their high structural consistency manifested by the strong conjugacy of latent axes directed towards the maximum data variability and the statistical proximity of the coordinates of river sites where the monitoring stations are located. Concurrently, the identified multiyear fluctuations in the taxonomic composition of hydrobionts make it possible to consider the ecosystem studied a complex of nonequilibrium dynamically changing species consortiums.

The interrelationship between the planktonic and bottom communities identified using multidimensional ordination techniques is determined both by biotic interactions and mutually consistent reactions to changes in aquatic environmental factors. It is established that the key hydrochemical factors significantly affecting the taxonomic structure include the total water hardness and the concentrations of phosphates, chlorides, dissolved oxygen, and ammonium ions. Species groups with different (higher or lower) degrees of tolerance to the above-listed abiotic impacts have been identified.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The authors declare that they have no conflict of interest.

Statement on the welfare of animals. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

REFERENCES

Dolédec, S. and Chessel, D., Co-inertia analysis: an alternative method for studying species-environment relationships, *Freshwater Biol.*, 1994, vol. 31, pp. 277−294.

Dray, S., Chessel, D., and Thioulouse, J., Procrustean co-inertia analysis for the linking of multi-variate datasets, *Écoscience*, 2003, vol. 10, o. 3, pp. 110−119.

González, I., Lê Cao, K.-A., Davis, M.J., and Déjean, S., Visualising associations between paired "omics" data sets, *BioData Minning*, 2012, vol. 5, p. 19.
https://doi.org/10.1186/1756-0381-5-19

Hervé, M., Aide-mémoire de statistique appliquée à la biologie. Construire son étude et analyser les résultats à l'aide du logiciel R., Togliatti, 2018. https://stok1946.blogspot.com. Accessed July 24, 2018.

Hervé, M.R., Nicolè, F., and Lê Cao, K.A., Multivariate analysis of multiple datasets: a practical guide for chemical ecology, *J. Chem. Ecol.*, 2018, vol. 44, no. 3, pp. 215−234.
https://doi.org/10.1007/s10886-018-0932-6

Lê Cao, K.-A., Rossouw, D., Robert-Granié, C., and Besse, P., A sparse PLS for variable selection when integrating omics data, *Stat. Appl. Genet. Mol. Biol.*, 2008, vol. 7, no. 1, p. 35.
https://doi.org/10.2202/1544-6115.1390

Shitikov, V.K. and Rozenberg, G.S., *Randomizatsiia i butstrep: statisticheskii analiz v biologii i ekologii s ispol'zovaniem R* (Randomization and Bootstrap: Statistical Analysis in Biology and Ecology using R), Togliatti: Kassandra, 2014.

Shitikov, V.K. and Zinchenko, T.D., Multivariate statistical analysis of ecological communities (review), *Theor. Probl. Ekol.*, 2019. no. 1, pp. 5−11.
https://doi.org/10.25750/1995-4301-2019-1-005-011

Singh, A., Gautier, B., Shannon, C.P., Vacher, M., Rohart, F., Tebutt, S.J., and Le Cao, K.-A., DIABLO−an integrative, multi-omics, multivariate method for multi-group classification. *BioRxiv*, 2016, no. 067611, pp. 1−50.
https://doi.org/10.1101/067611

Wold, H., Partial least squares, in *Encyclopedia of Statistical Sciences*, Kotz, S. and Johnson, N., Eds., New York: Wiley, 1985, pp. 581−591.

Zinchenko, T.D., Golovatjuk, L.V., Vykhristjuk, L.A., and Shitikov, V.K., Diversity and structure of macrozoobenthos communities in the highly mineralized Khara River (near Elton Lake), *Povolzh. Ekol. Zh.*, 2010, no. 1, pp. 14−30.

Zinchenko, T.D., Golovatyuk, L.V., Abrosimova, E.V., and Popchenko, T.V., Macrozoobenthos in saline rivers in the Lake Elton basin: spatial and temporal dynamics, *Inland Water Biol.*, 2017, vol. 10, no. 4, pp. 384−398.

Zinchenko, T.D., Shitikov, V.K., Golovatyuk, L.V., Gusakov, V.A., and Lazareva, V.I., Plankton and bottom communities in the saline rivers of Lake Elton Basin: statistical analysis of dependences, *Arid Ecosyst.*, 2018, vol. 8, no. 3, pp. 225−230.

Zinchenko, T.D., Shitikov, V.K., Golovatyuk, L.V., Gusakov, V.A., and Lazareva, V.I., Analysis of relations between communities of hydrobionts in saline rivers by multidimensional block ordination, *Inland Water Biol.*, 2019, vol. 12, no. 2, pp. 104−110.

Translated by L. Emeliyanov

SPELL: 1. OK