LIMNOLOGICA

Verification of the dualism ordering method by the canonical correspondence analysis: fish community samples

Tadeusz Penczak[1,*], Francisco Godinho[2], Angelo A. Agostinho[3]

[1] Department of Ecology and Vertebrate Zoology, University of Łódź, Poland
[2] Estaçáo Florestal Nacional, Posto Apicola Tapada da Ajuda Lisboa, Portugal
[3] NUPELIA, Universidade Estadual de Maringa, Maringa, PR, Brazil

Abstract

Two ordination techniques (Romaniszyn diagram, RD, and Canonical correspondence analysis, CCA) were applied to the same data set (fish assemblage composition along two Paraná river tributaries and associated environmental variables), to identify spatial community patterns and further to evaluate the performance of RD. The major patterns in these Brazilian streams were successfully revealed by both methods (i.e. clear separation of both tributaries on the basis of fish assemblage composition and environmental factors), but more complex situations may limit the application of RD in comparison with CCA.

Key words: Canonical correspondence analysis – Romaniszyn diagram – conformable results – Paraná River tributaries – fish assemblages

Introduction

Canonical correspondence analysis (CCA) proposed by TER BRAAK (1986) was successfully used to investigate community structure and its underlying environmental basis, including several fish studies (e.g. COPP 1992; TAYLOR et al. 1993; WILLIAMS et al. 1996; GODINHO et al. 1997; RODRÍGUEZ & LEWIS 1997; GODINHO & FERREIRA 1998; GODINHO et al. 1998; TEJERINA-GARRO et al. 1998; PIRES et al. 1999). The principle of the dualism method and the synthetic diagram presenting its results (Romaniszyn diagramm, RD) that were developed by ROMANISZYN (1970) were applied a few times to investigate fish assemblages, but the authors or co-authors of these papers, with one exception (BALON & STEWART 1983), were Polish researchers (PENCZAK 1972; WITKOWSKI 1984; PENCZAK et al. 1991, 1994; ZALEWSKI et al. 1990; PRZYBYLSKI et al. 1991). As they published mainly in Polish journals, the RD method attracted no attention of new investigators.

Scientists are continuously looking for new and better methods for analysing community structure on the background of abiotic and biotic factors. According to MATTHEWS (1998) it is desirable to test the results from several analyses to determine if fish assemblages occur in distinct groups or if they occur randomly in multivariate space. Besides, comparing two methods can be helpful because when their results are similar, or better identical, they calibrate each other and increase the credibility and validity of both. Hence, the aim of this study is to give some added confidence and encourage researchers to use of less known Romaniszyn method. Also, we want to show that it is easy to interpret the Romaniszyn synthetic diagram, however the method is not explaining variance in species data. For a respective comparison we used fish assemblages from two small tributaries of the Paraná River for which quantitative fish samples and routine site characteristics were available (PENCZAK et al. 1994; AGOSTINHO & PENCZAK 1995).

*Corresponding author: Prof. Dr. Tadeusz Penczak, Department of Ecology and Vertebrate Zoology, University of Łódź, 12/16 Banacha Str., 90–237 Łódź, Poland; e-mail: penczakt@biol.uni.lodz.pl

Study area, Material and Methods

Eleven sites were located in two small, left bank tributaries of the Paraná River in the north-western corner of the Paraná State, Brazil. The Caracu River, 6.8 km long, empties directly to the Paraná, and the Água do Rancho River, 4 km long, empties to the Areia Branca River at 7.8 km from its outlet to the main corridor (Fig. 1). Valleys geology, soil, site morphology and some water parameters are described elsewhere (PENCZAK et al. 1994; AGOSTINHO & PENCZAK 1995).

A total of 1260 fish was collected and identified into 27 taxa; their affiliations to 14 families are presented in PENCZAK et al. (1994). Three successive catches at each site were undertaken by way of catch per constant unit effort electrofishing (CPUE), blocked between two nets (Table 1). Samples were collected on 28–31 October 1992. Two people waded upstream of 1.9–3.6 m wide streams with anode-dipnets using full-wave rectified current from a 3 kW generator with an output current of 220 V, and 1.5–2.5 A (PENCZAK et al. 1994; AGOSTINHO & PENCZAK 1995). In the Agua do Rancho Stream, whose water displayed a lower conductivity, several plastic perforated bags filled up with salt were sunk a

few meters upstream of the beginning of the site (PENCZAK et al. 1997).

Canonical correspondence analysis is an eigenvalue technique with which both biological data (i.e. species abundance) and environmental data, measured at a number of sites, are analysed. A diagram can be generated from the results to display simultaneously the similarities between sites, based on species associations, and the relationships between the abundance of different species and environmental variables (TER BRAAK 1986, 1990; TER BRAAK & VERDONSCHOT 1995). The algorithm used in CCA is reciprocal averaging (as in correspondence analysis) of the number of species × sites matrix together with linear least-squares regression on the environmental variables. For analysis two matrices were assembled: – fish taxa abundance by sites and environmental variables by sites. Rare taxa (i.e. occurring in less than 3 sites) were removed to prevent analysis distortion and both abundance data and environmental variables were not transformed. First, a general CCA was performed with all the environmental variables and second, another CCA was made only with the environmental variables selected by a forward selection procedure available in CANOCO (version > 3.1). This procedure aims at
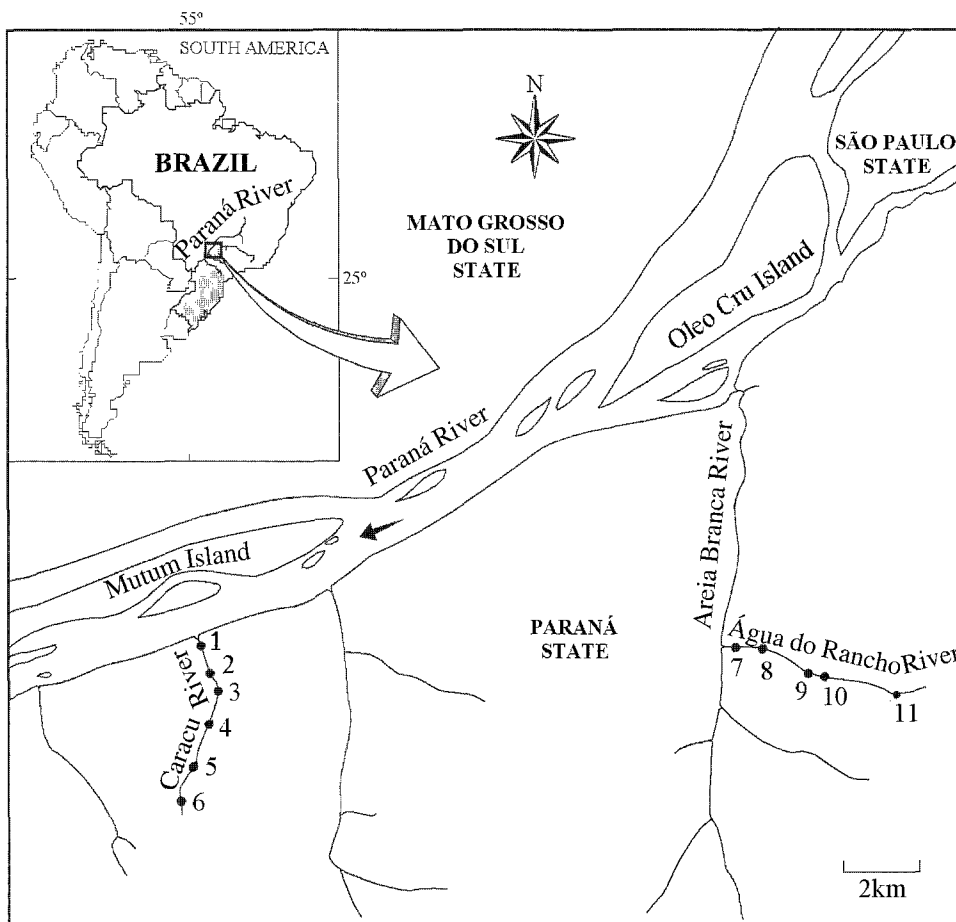

Fig. 1. Study area with sites indicated.

selecting the best group of variables that explain the patterns almost as well as the full set of environmental variables, with variables being selected by decreasing additional explanatory power (TER BRAAK & VERDONSCHOT 1995). For both CCAs, a Monte Carlo simulation with 1000 permutations was used to test the significance of both the first axis eigenvalue and trace (the sum of all eigenvalues) (TER BRAAK 1987, 1990).

The analysis proposed by ROMANISZYN (1970) was printed in Polish. However, it was described and exemplified in English in a few papers, and in detail by PENCZAK et al. (2000). The method is based on the principle of dualism, consisting in ordering data twice using the similarity definition: first similarity is calculated between each pair of sites ('columns' in Table 1), and second between each pair of species ('rows' in Table 1). The similarity (s) is calculated using the MARCZEWSKI & STEINHAUS (1958) equation:

$$s = w/a + b - w * 100,$$

where s is similarity of two collections, w is the total of the lower number of specimens of each pair of species common for two given analysed collections, a is the total of specimens of a species at the site (or the total number of sites with species A), and b is the total number of specimens of a species in site (or the total number of sites with species B). Using similarities calculated between 'columns' in one table and afterwards 'rows' in another table, two branched and two-dimensional dendrites are constructed. They are then converted to linear ones (i.e. ordination axis), respecting the principle that the weakest connection must be 'broken' to insert between sites/species those that are not in linear order (see examples in PENCZAK et al. 1994, 2000). Clusters of sites and species can be distinguished by quotients calculated between neighbouring, decreasing values of similarity; of course, separately for sites and species. If a given quotient occurring in the middle of a quotient sequence differs from its neighbouring ones, then its dividend (similarity value) constitutes the lower bound of a similarity for defining a cluster. All clusters below that value of similarity are considered separately (ROMANISZYN 1970).

Results and Discussion

Romaniszyn's synthetic diagram (Fig. 2) and CCA (Fig. 4), both clearly separated sites belonging to the Caracu and Água do Rancho Rivers, despite their spatial proximity. Fish assemblages are considerably dissimilar between the two streams. For example, taxa such as *Phenacorhamdia* sp., *Cetopsorhamdia iheringi*, *Bryconamericus stramineus*, *Astyanax scabripinis* and *Microlepidogaster* sp. were only found in the Água do Rancho River whereas *Astyanax bimaculatus*, *Characidium*

Table 1. Number of fish collected by successive electric fishing and stop net fishing in the Caracu River (sites 1–6) and in the Água do Rancho River (sites 7–11) used for CCA; species represented by one specimen in one site were omitted (PENCZAK et al. 1994).

| Scientific name	Caracu River						Agua do Rancho River				
	1	2	3	4	5	6	7	8	9	10	11
Astyanax scabripinis (Eigenmann, 1927)							5	25	16	18	93
Bryconamericus stramineus (Eigenmann, 1908)							14	18	32	13	
Astyanax bimaculatus (Linnaeus, 1758)	8	9	90	15	3	3		1			
Characidium fasciatum Reinhardt, 1866	12	4	2					1			
Cheirodon notomelas (Eigenmann, 1915)	1	3			2						
Rhamdia quelen (Quoy & Gaimard, 1824)							6	21	1		13
Nannorhamdia schubarti Gomes, 1956	2	7						1	1		
Cetopsorhamdia iheringi Schubart & Gomes, 1959							9	9	11	9	
Phenacorhamdia sp.							11		5	12	
Hypostomus ancistroides Ihering, 1911	27	25	21	9	3	3	50	48	60	48	2
Microlepidogaster sp.							169	48	2	7	1
Leporinus silvestris Boulanger, 1902		2	5	1							
Callichthys callichthys (Linnaeus, 1758)	1	2		8							
Phalocerus caudimaculatus (Hensel, 1868)				7	1	30		32			19
Synbranchus marmoratus (Bloch, 1795)			1			4	1	1	1	2	4
Eigenmania trilineata (Lopez & Castello, 1966)							8	1			
Gymnotus carapo (Linnaeus, 1758)	11	11	5	12	1	3	5	19	2	4	10
Hoplias malabaricus (Bloch, 1794)											
Steindachnerina insculpta (Fernandes-Yepez, 1948)	3										
Prochilodus lineatus Steindachner, 1882			4								
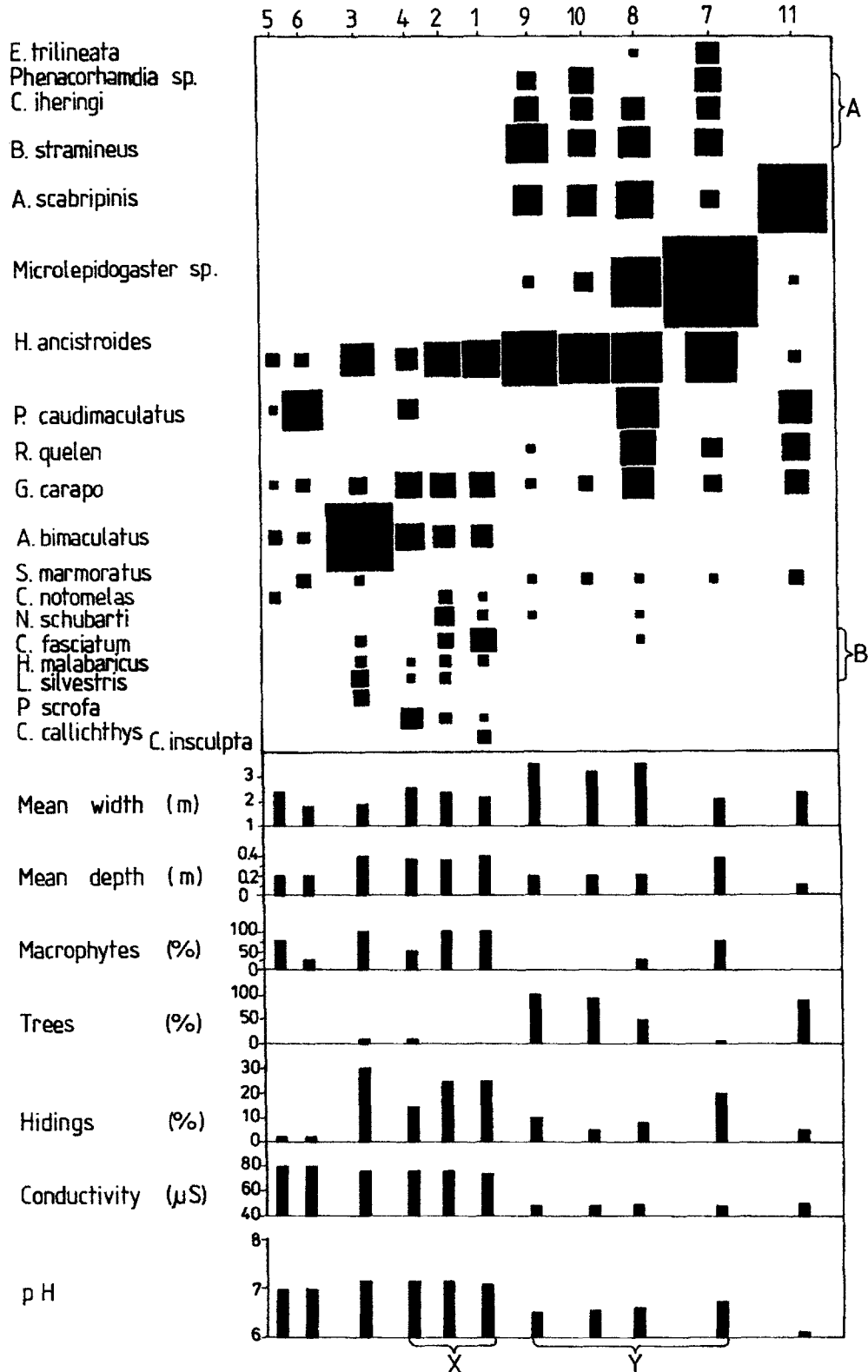
Fig. 2. Synthetic diagram showing quantitative species distribution in relation to sites (1–11). Sites characteristics related with fish populations abundance are presented below the diagram. A and B are species clusters; X and Y are site clusters - distinguished on the basis of a quotient calculated between neighbouring similarity values (after PENCZAK et al. 1994); in the present publication *Steindachnerina insculpta* was named *Curimata insculpta*, and *Prochilodus lineatus* was named *Prochilodus scrofa*.
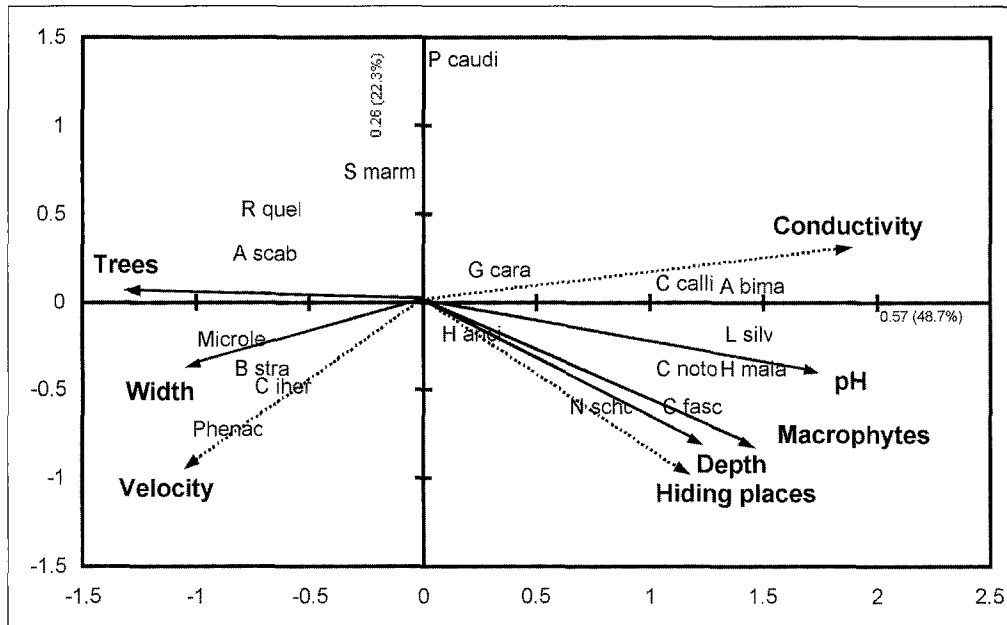
Fig. 3. Diagram of axis one and two for the Canonical correspondence analysis relating fish taxa abundance and environmental variables assessed in 11 sites in two Parana tributaries. Taxa are depicted by abbreviations of scientific names (compare Fig. 2). Environmental variables are represented by arrows that roughly point towards the factor direction of maximum variation. The length of an arrow is relative to the importance of that variable in assemblage ordination. Eigenvalues and the % of variation in species abundance (between brackets) explained by each axis is also presented. Variables selected by a forward selection procedure available in CANOCO are depicted by a dotted line. The biplot resulting from the CCA performed with this procedure was almost identical to the biplot of the global CCA and is not presented.
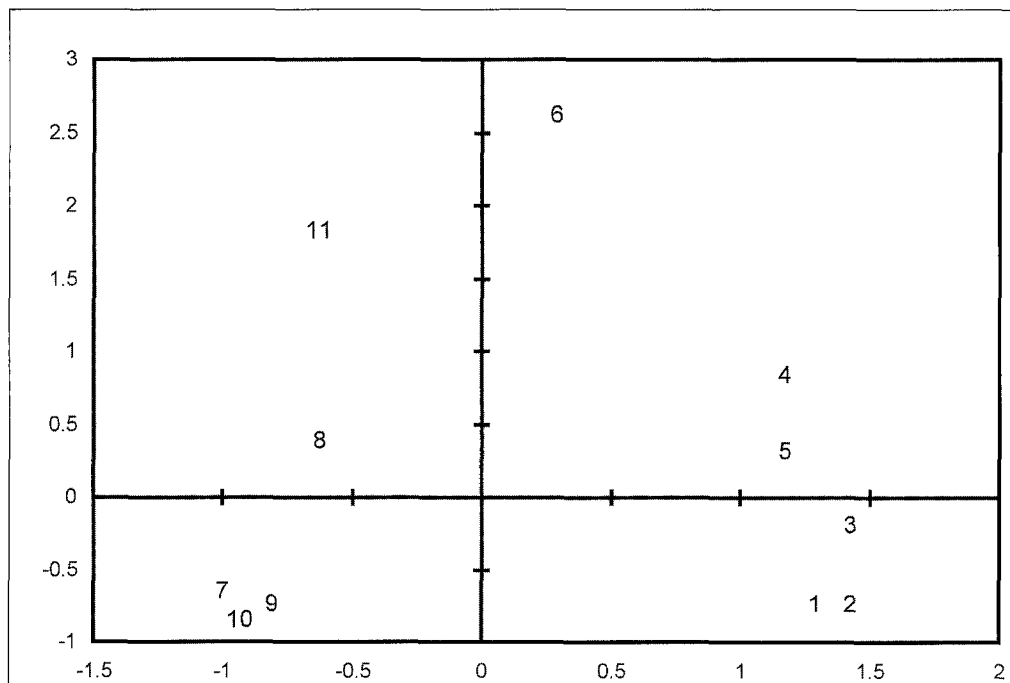

Fig. 4. Diagram of axis one and two for the Canonical correspondence analysis relating site scores of the Caracu and Água do Rancho Rivers.

fasciatum, Hoplias malabaricus, Leporinus silvestris and *Callichthys callichthys* were only collected in the Caracu River. The synthetic diagram separated sites from the two streams and the species were placed on a gradient: – species exclusively occurring in the Água do Rancho; – species occurring in both streams; – species occurring exclusively in the Caracu (Fig. 2). The species of the two streams are located along the first axis of CCA (Fig. 3), which account for the greatest amount of variation, from the fish fauna characterising the Água do Rancho on the left of the diagram to that of the Caracu on the right. Species occurring in both streams were placed in the middle of the first axis. In this way the assemblages are partially separated; the assemblages overlap only in that area where species common to both streams occur.

According to the environmental variables it can be seen that the clear differences in assemblage composition between the two streams were associated with the also clear differences in environmental characteristics between the studied streams. In RD, as in other indirect ordination methods, the analysis of environmental gradients is performed a posteriori, after ordination of sites and species. From Fig. 2 it is apparent that sites and corresponding assemblages from the Caracu River were associated with higher conductivity, higher pH, and, on average, more refuges, more macrophyte cover, greater depth, smaller width, lower velocities and less trees along the banks (Fig. 2). In contrast, sites in the Água do Rancho (and corresponding assemblages) were related to the opposite environmental conditions.

In total, CCA accounted for 92.1% of the variation in fish assemblage composition (i.e. the quotient between the trace and the sum of all eigenvalues of a correspondence analysis of the species × sites matrix, TER BRŘAK 1987), with the first and second axis accounting for 70.0% of variation in species abundance. The analysis was also significant ($P < 0.01$, for both the first axis and trace). Such real differences between the investigated streams and their fish faunas decided of their clear separation using RD.

To the best of our knowledge, this is the highest variation explained by CCA in a fish study, followed by WILLIAMS et al. (1996), accounting for as much as 71% of total variation in assemblage composition (significant at $P < 0.01$). In fact, conductivity alone (the major predictor of fish assemblage composition selected in the forward selection procedure) accounts for a large portion of the biotic variation (41%) since it clearly separates both streams; conductivity plus the two other variables selected in the forward selection accounted for 71% of biotic variation (Fig. 3). Other variables also nonetheless determined to be important (and all variables included in the global CCA were significantly correlated with the canonical axes, $P < 0.05$) in species ordi-nation were correlated with conductivity and thus the separation of assemblages and sites from the two streams obtained with just conductivity may have reflected differences in other variables (note that the diagrams for both CCA and RD are very similar in the distribution of sites and species).

Acknowledgements

References

AGOSTINHO, A.A. & PENCZAK, T. (1995): Populations and production of fish in two small tributaries of the Parana River, Parana, Brazil. Hydrobiologia **312**: 153–166.

BALON, E.K. & STEWART, D.J. (1983): Fish assemblage in a river with unusual gradient (Luongo, Africa - Zaire system), reflections on river zonation, and description of another new species. Env. Biol. Fish. **9**: 225–252.

COPP, G.H. (1992): An empirical model for predicting microhabitat of 0+ juvenile fishes in a lowland river catchment. Oecologia **91**: 338–345.

GODINHO, F.N. & FERREIRA, M.T. (1998): The relative influences of exotic species and environmental factors on an Iberian native fish community. Env. Biol. Fish. **51**: 41–51.

– FERREIRA, M.T. & CORTES, R.V. (1997): Composition and spatial organization of fish assemblages in the lower Guadiana basin, southern Iberia. Ecol. Freshw. Fish **6**: 134–143.

– FERREIRA, M.T. & PORTUGAL E CASTRO, M.I. (1998): Fish assemblage composition in relation to environmental gradients in Portuguese reservoirs. Aquat. Living Resour. **11**: 325–334.

MARCZEWSKI, E. & STEINHAUS, H. (1958): On a certain distance of sets and the corresponding distance of function. Coll. Math. **6**: 319–327.

MATTHEWS, W.J. (1998): Patterns in freshwater fish ecology. Kluwer Academic Publishers. New York, 756 pp.

PENCZAK, T. (1972): Structure of fish groupings in the rivers and streams of the Nida drainage basin. Ekol. Pol. **20**: 327–344.

– PENCZAK, T., AGOSTINHO, A.A. & HAHN, N.S. (2000): Applying an ordination technique for fish diet comparison. Brazil. Arch. Biol. Tech. **43**: 101–110.

– AGOSTINHO, A.A. & OKADA, E.K. (1994): Fish diversity and community structure in two small tributaries of the Parana River, Parana State, Brazil. Hydrobiologia **294**: 243–251.

– AGOSTINHO, A.A., GŁOWACKI, L. & GOMES, L.C. (1997): The effect of artificial increases in water conductivity on the efficiency of electric fishing in tropical streams (Paraná, Brazil). Hydrobiologia **350**: 189–201.

– Forbes, I., Coles T.F., Atkin, T. & Hill, T. (1991): Fish community structure in the rivers of Lincolnshire and South Humberside, England. Hydrobiologia **211**: 1–9.

Pires, A.M., Cowx, I.G. & Coelho, M.M. (1999): Seasonal change in fish community structure of intermittent streams in the middle reaches of the Guadiana basin, Portugal. J. Fish Biol. **54**: 235–249.

Przybylski, M., Biro, P., Zalewski, M., Tatrai, I. & Frankiewicz, P. (1991): The structure of fish communities in streams of the northern part of the catchment area of Lake Balaton (Hungary). Acta Hydrobiol. **33**: 135–148.

Rodriguez, M.A. & Lewis, Jr. W.M. (1997): Structure of fish assemblages along environmental gradients in floodplain lakes of the Orinoco River. Ecol. Monogr. **67**: 109–128.

Romaniszyn, W. (1970): Próba interpretacji tendencji skupiskowych zwierząt w oparciu o definicję podobieństwa i odległości [An attempt at interpreting agglomerative tendencies of animals based on definition of similarity and distance]. Wiad. Ekol. **16**: 306–327 (English summary).

Taylor, C.M., Winston, M.R. & Matthews, W.J. (1993): Fish species-environment abundance relationships in a Great Plains river system. Ecography **16**: 16–33.

Ter Braak, C.J.F. (1986): Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **67**: 1167–1179.

– (1987): Canoco – A fortran program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis (version 2.1). Wageningen, Netherlands: ITI-TNO Institute of Applied Computer Sciences.

– (1990): CANOCO version 3.1. Update notes. Wageningen, Netherlands: Agricultural Mathematics Group.

– & Verdonschot, P.F.M. (1995): Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquat. Sci. **57**: 255–289.

Tejerina-Garro, F.L., Fortin, R. & Rodriguez, M.A. (1998): Fish community structure in relation to environmental variation in floodplain lakes of the Araguaira River, Amazon Basin. Env. Biol. Fish. **51**: 399–410.

Williams, L.R., Toepfer, C.S. & Martinez, A.D. (1996): The relationships between fish assemblages and environmental gradients in an Oklahoma prairie stream. J. Freshwater Ecol. **11**: 459–468.

Witkowski, A. (1984): Structure of communities and biomass of ichthyofauna in the Biebrza River, its old river beds and affluents. Pol. Ecol. Stud. **10**: 447–474.

Zalewski, M., Frankiewicz, P., Przybylski, M., Banbura, J. & Nowak, M. (1990): Structure and dynamics of fish communities in temperate rivers in relation to the abiotic-biotic regulatory continuum concept. Pol. Arch. Hydrobiol. **37**: 151–176.