Hydrobiologia **263**: 109–118, 1993.

Macroinvertebrate distribution in streams: a comparison of CA ordination with biotic indices

Bruno Rossaro & Annamaria Pietrangelo
University of L'Aquila, Dipartimento di Scienze Ambientali, Coppito, I-67010 Coppito L'Aquila, Italy

Received 24 June 1992; in revised form 5 January 1993; accepted 14 January 1993

Key words: macroinvertebrates, ordination methods, biotic index, running waters

Abstract

Macroinvertebrates were collected in running waters in Italy, analyzed with correspondence analysis (CA) and with the calculation of 8 biotic indices. Then the CA ordination axes were correlated with 19 environmental variables and with biotic indices.

The first CA axis is easily interpreted as an upstream-downstream gradient and is correlated with physical factors (particle size, slope etc.), whereas the second axis separated permanent waters from temporary ones.

The first CA axis correlated with many biotic indices suggesting that biotic indices are strongly influenced by physical factors. Multiple regressions with 2 biotic indices as criterion and the 19 environmental factors as predictor variables confirm the importance of physical factors in determining the values of the biotic indices.

The advantages and drawbacks of the use of CA instead of biotic indices is discussed.

Introduction

Water quality indices have been formulated to provide a routine technique for use in water monitoring. Macroinvertebrates are considered good indicators of environmental pollution in running waters and are extensively used in the formulation of biotic indices. It is generally accepted that macrofauna respond both to hydraulic, organic, and toxic stress with reduction of sensitive species and proliferation of tolerant ones (Verdonschot, 1990).

Nevertheless, little emphasis has been given to the influence of natural factors in determining macroinvertebrates species structure and, subsequently, the biotic index value.

For example, significant variations in diversity and number of taxa were observed in streams of different order (Crunkilton & Duchrow, 1991).

It must be emphasized that knowledge about the response of taxa to different factors is needed in order to be able to separate the effects of pollution from the effects of natural variables that affect community structure.

Many authors have analyzed factors responsible for establishment, maintenance and modification of benthic invertebrate communities (Minshall & Petersen, 1985). Both spatial and temporal variation were examined. Different substrates that determined spatial separation were found to be more important than seasonal separation in some studies (Verdonschot, *op. cit.*), whereas a separation of samples by season was emphasized in others (Hilsenhoff, 1988; Miller & Stout, 1989).

Multivariate ordination methods have been extensively used by running water ecologists to analyze the response of macroinvertebrates.

Verneaux (1973) used correspondence analysis to analyze streams in the Doubs river catchment and emphasized that taxa and sites plotted on the first two axes were ordered according to an U-shaped curve that described different community types along an upstream-downstream gradient. Polluted stations were plotted outside the U curve towards the center of the plot. This was interpreted as the consequence of the presence of tolerant ubiquitous species in the polluted stations.

Sheldon & Haick (1981) used principal component analysis and canonical correlation analysis. Three habitat – fauna interactions were identified. The first corresponded to a gradient from eroding to depositing substrates, the second distinguished areas of fine from leafy detritus, the third included velocity – stream size interactions.

Verdonschot (*op. cit.*) used detrended constrained canonical ordination (DCCA) and observed that in streams the most important gradient that separates taxa is a transition from natural to regulated and/or organic polluted streams. The second gradient separated temporary from permanent waters.

Current knowledge emphasizes that the use of macroinvertebrates to develop water quality criteria is hindered by the interaction between the influence of natural and anthropogenic factors. For example organic pollution and river regulation (hydraulic stress) interact with organic debris content, water speed, etc. Polluted stations are often downstream stations, in these cases changes in community structure can be due both to changes in community type along the upstream-downstream gradient and changes in the degree of pollution.

At present the opinion is that biological measures of water quality (biotic indices) estimate both the effect of pollution and the effect of natural variables (Armitage *et al.*, 1983; Moss *et al.*, 1987).

There is the need to continue computerization of biological data (list of taxa, presence in different sites) and environmental data (river type, current speed, nature of the bottom) and to explore the relation between the environmental factors and the presence of organisms by multivariate analysis (De Pauw & Vanhooren, 1983). This is required for different geographic areas, water types and sources of pollution.

The aim of this study is to investigate the capability of the correspondence analysis (CA) (Ter Braak & Prentice, 1988) to summarize data on running water macroinvertebrates and to compare CA ordination with different biotic indices.

Material and methods

Samples were collected from different sites in streams in Italy (Fig. 1) with a hand net according to the standard methods discussed in Ghetti (1986).

The sites are distributed from headwaters to large streams and rivers (Ticino, Trebbia, Secchia, Arno).

Nineteen environmental variables were measured, mean values and standard deviations are listed in Table 1. Particle size, heterogeneity of particle size, current speed and visual index were filed as classified variables. Ordinal scale values are in Table 2.

Macroinvertebrates were determined to the taxonomic level established for the calculation of EBI (Ghetti, 1986).

Input data were presence – absence of 160 taxa in 651 sites. The list of taxa is available and not given for reason of space.

Different biotic indices were calculated:

1 CQ = water quality class (Ghetti, 1986)
2 EBI = extended biotic index (Ghetti, 1986)
3 BMWP = biological monitoring working party (Metcalfe, 1989)
4 ASPT = average score per taxon (Metcalfe, 1989)
5 US = total number of taxonomic units at the taxonomic level of EBI

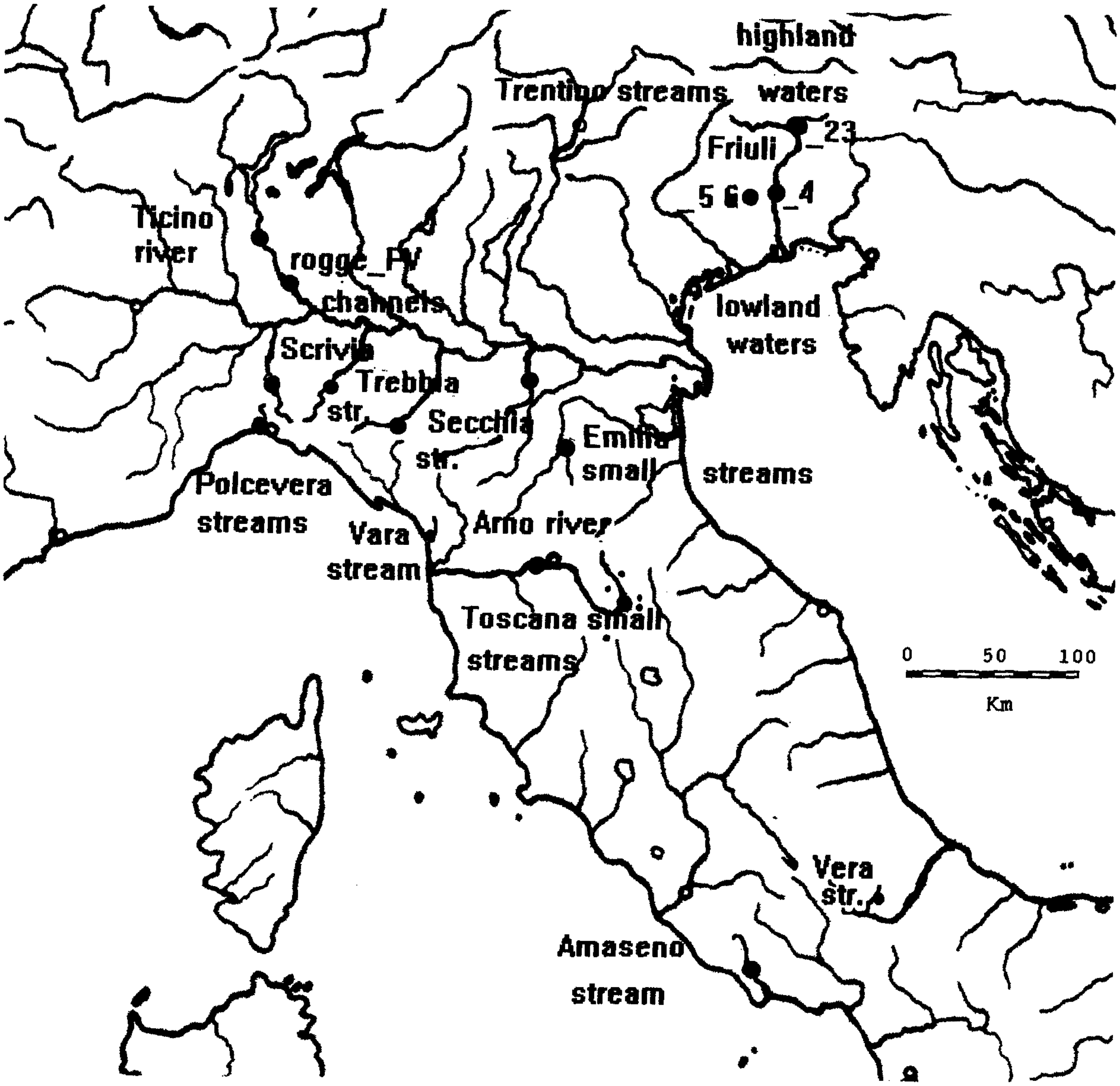

Fig. 1. Map of the sampled sites.

6 N-BMWP = total number of taxonomic units at the taxonomic level required by BMWP
7 IRE = extended ratio index (Stoch, 1986)
8 Hindd = Shannon diversity index.

Diversity was calculated whenever quantitative data were available (102 samples over 651).

Not all environmental variables were measured at all sites. The number of measurement available is given in Table 1.

Indirect gradient analysis (correspondence analysis, CA) was performed without detrending, because there is no agreement about the advantages of detrending (Wartenberg *et al.*, 1987). Downweighting of rare species was applied. The data analysis was accomplished using all 651 sites and 160 taxa together.

The results of the ordination is summarized by the eigenvalue, that is a measure of the goodness of separation in the distribution of taxa along an

Table 1. Mean and standard deviation of 19 environmental variables and number of observations available for each variable, C.V. = classified variables.

	Unit of measure	Mean	Stand dev.	No obs.
Particle size	C.V.	15.47	4.94	440
Part. size heterog.	C.V.	1.91	0.67	440
Slope	‰	4.66	2.40	152
Source distance	km	31.73	47.78	580
River catchment	km^2	313.15	160.88	137
River discharge	$m^3 s^{-1}$	4.66	5.40	236
River disch. summer	$m^3 s^{-1}$	4.01	1.40	56
River disch. at outlet	$m^3 s^{-1}$	30.11	15.52	135
Current speed	C.V.	3.79	0.56	181
Total length	km	121.10	53.86	202
Wetted stream wide	m	26.12	30.68	289
Dry stream wide	m	45.16	22.97	108
Depth	m	0.76	0.46	217
Water temperature	°C	15.58	2.05	88
Canopy cover	%	39.24	15.98	135
Conductivity (EC)	μS	232.67	45.25	92
BOD_5	$mg\,l^{-1}$	4.74	3.24	123
Equivalent habitants	$ind\,km^{-2}$	431.13	453.32	189
Visual index	C.V.	4.94	0.78	225

ordination axis. The sum of eigenvalues is a measure of all biological variation present on all axes.

The correlation coefficient between ordination axes, environmental variables (Table 3) and biotic indices (Table 4) were calculated, using the NTSYS program (Rohlf, 1985) that accepts missing data.

Two multiple regression analyses with environmental data as predictor variables and EBI or ASPT as criterion variables were then carried out.

Results

The eigenvalue in CA is large for the first axis (0.35) and much lower for the second axis (0.19) (Table 3). This indicates that there is a strong gradient along the first axis and a much weaker one along the second axis. The first ordination axis can be easily interpreted as an upstream – downstream gradient. It is inversely correlated with particle size, slope, and directly with conductivity (EC) (Table 3 and Fig. 2). It can be interpreted as an upstream-downstream gradient.

Table 2. Codes used in classified variables. Substrate heterogeneity was coded in the following manner: when only one fraction was present it was coded 1, when more than one fraction was present (for example gravel + coarse sand) it was coded as the number of fractions (2 means two fractions present)

Particle size	
Boulder (> 256 mm)	7
Cobble (64–265 mm)	6
Pebble (64–16 mm)	5
Gravel (16–2 mm)	4
Coarse sand (2–0.25 mm)	3
Fine sand (0.25 mm–62.5 μm)	2
Silt (62.5–3.9 μm)	1
Clay (< 3.9 μm)	0
Visual index	
Clean water	8
Presence of macrophytes	7
Periphyton	6
Bare bottom	5
Turbid water	4
Pres. of organic matter	3.5
Vegetable debris	3.5
Trace of anaerobiosis	3
Presence of hydrocarbons	2
Presence of surfactants	1
Current speed	
Low	2
Low to moderate	3
Moderate	4
Moderate to high	5
High (> 0.8 ms^{-1})	6

The second axis separates large permanent water bodies from small temporary water bodies. It is inversely correlated with slope, total river length and visual index. The correlation coefficients between environmental variables and the 1st ordination axis are the abscissae and the correlation coefficients between environmental variables and the 2nd axis are the ordinates in Fig. 2.

An ordination diagram for axes 1 (horizontally) and 2 (vertically) can be drawn for taxa (Fig. 3) and for sites (Fig. 4).

Taxa belonging to different groups are separated in the diagram. Most genera of Plecoptera

Table 3. Correlation coefficients between environmental variables and ordination axes in correspondence analysis (CA).

	CA Axis-1	CA Axis-2
Eigenvalue	0.35	0.19
Particle size	- 0.53	- 0.12
Part. size heterog.	- 0.12	- 0.01
Slope	- 0.51	- 0.50
Source distance	0.13	- 0.01
River catchment	- 0.34	- 0.34
River discharge	- 0.03	- 0.08
River disch. summer	- 0.47	- 0.43
River disch. at outlet	- 0.32	- 0.39
Current speed	- 0.18	- 0.14
Total length	- 0.11	- 0.57
Wetted stream wide	0.06	- 0.01
Dry stream wide	0.23	0.14
Depth	0.31	- 0.26
Water temp.	0.21	0.29
Canopy cover	0.17	- 0.14
EC	0.70	0.35
BOD	0.27	0.25
Equiv. habit.	- 0.06	- 0.12
Visual index	- 0.44	- 0.54

are situated in the left part of the figure (high negative loadings in the first axis), many Ephemeroptera genera are also in the left part, but with more scatter, whereas Hirudinea, Odonata and Hemiptera appear on the right part (high positive loadings in the first axis). A substitution of taxa belonging to different Insects orders along the first axis is evident in the following sequence: Plecoptera – Ephemeroptera – Trichoptera – Hemiptera. Diptera and Coleoptera span over all length of the first axis.

The second axis separates Odonata, Hemiptera, Dipera and Hirudinea in the upper part of the figure from Triclads, Crustacea, Mollusca and Trichoptera, which are plotted in the lower part.

Different taxa belonging to the same taxonomic group (families of the same order for example) are often scattered (this cannot be seen in Fig. 3, that gives mean values) and this is in agreement with their ecological response. Within the order of Diptera the rheobiont family Blephariceridae has very high negative scores in the first axis and it is situated in the left, whereas Ephydridae, Syrphidae have high loadings and appear in the right-hand side of Fig. 3.

Sites from cold stony bottom stream in Alps (Trentino, Friuli-23) are crowded in the left part of Fig. 4 (high negative loadings), sites from lowland waters (Rogge-PV) are on the right, sites from large rivers (Ticino) and from lowland permanent waters (Friuli-4, Friuli-56) are in the lower right part of Fig. 4. Small streams subject to drying in summer (streams situated in Piacenza and Reggio Emilia provinces, in Padana lowland territory) are in the upper right part of the diagram.

In Fig. 4 sites are ordered from left to right according to an upstream–downstream gradient.

Large rivers (Ticino) and lowland permanent waters make a large cluster separated from the major upstream-downstream gradient in the lower right part of the diagram.

Polluted sites often appear as outliers in relation to a cluster of sites belonging to the same

Table 4. Correlation coefficients between biotic indices and ordination.

Axes	Axis-1	Axis-2	Axis-3	EBI	CQ	BMWP	ASPT	US	N-BMWP
EBI	- 0.52	- 0.43	- 0.29	1.00					
CQ	0.04	0.12	0.11	- 0.62	1.00				
BMWP	- 0.44	- 0.38	- 0.19	0.89	- 0.46	1.00			
ASPT	- 0.75	- 0.37	- 0.33	0.77	- 0.24	0.71	1.00		
US	- 0.29	- 0.35	- 0.10	0.85	- 0.50	0.94	0.54	1.00	
N-BMWP	- 0.18	- 0.35	- 0.15	0.81	- 0.49	0.93	0.49	0.94	1.00
Hindd	- 0.53	0.05	0.01	0.79	m0.27	0.51	0.69	0.45	0.44
IRE	0.47	- 0.45	- 0.21	0.78	- 0.63	- 0.08	- 0.25	0.10	0.08

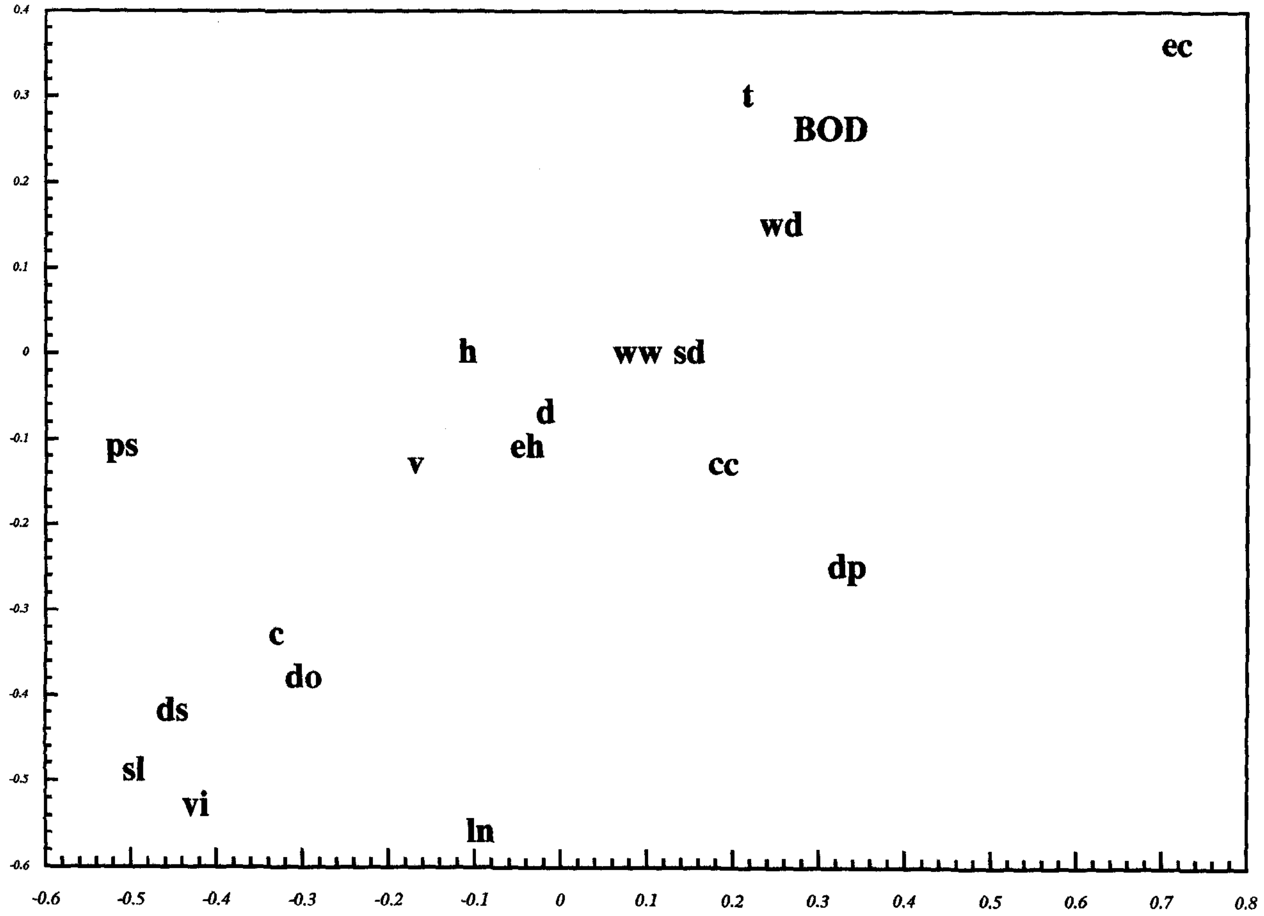

Fig. 2. Correlations between environmental variables and the first 2 CA axes; abscissa: correlations with the 1st axis, ordinate: correlations with the 2nd axis.

Symbols abbreviations (see also Table 1):
c = Size of catchment basin
sd = Source distance
wd = Dry stream wide
ww = Wetted stream wide
dp = mean river depth at the sampled site
sl = Stream slope
ln = Total river length
d = River discharge measured at the site
do = River discharge measured at river outlet
ds = River discharge measured in summer
v = Current speed
ps = Particle size of substrate (Table 2)
h = Particle size heterogeneity
t = Water temperature
BOD = Biochemical oxygen demand
ec = Electric conductivity (EC)
eh = Number of equivalent habitant
vi = See Table 2
cc = % of canopy cover

stream and many polluted stations are also downstream stations.

Correlation coefficients between biotic indices and ordination axes of CA show that all these variables are correlated with each other (Table 4). ASPT is the index more correlated with the first ordination axis, but EBI and BMWP are also well correlated.

To analyze the performance of the biotic indices, two multiple regressions were carried out with EBI and ASPT as criterion and environmental variables as predictors. Both analyses gave a multiple correlation coefficient ≈ 0.53 (Table 5), the variance ratio test of the hypothesis of zero relationship between criterion and predictor environmental variables (Cooley & Lohnes, 1973) is

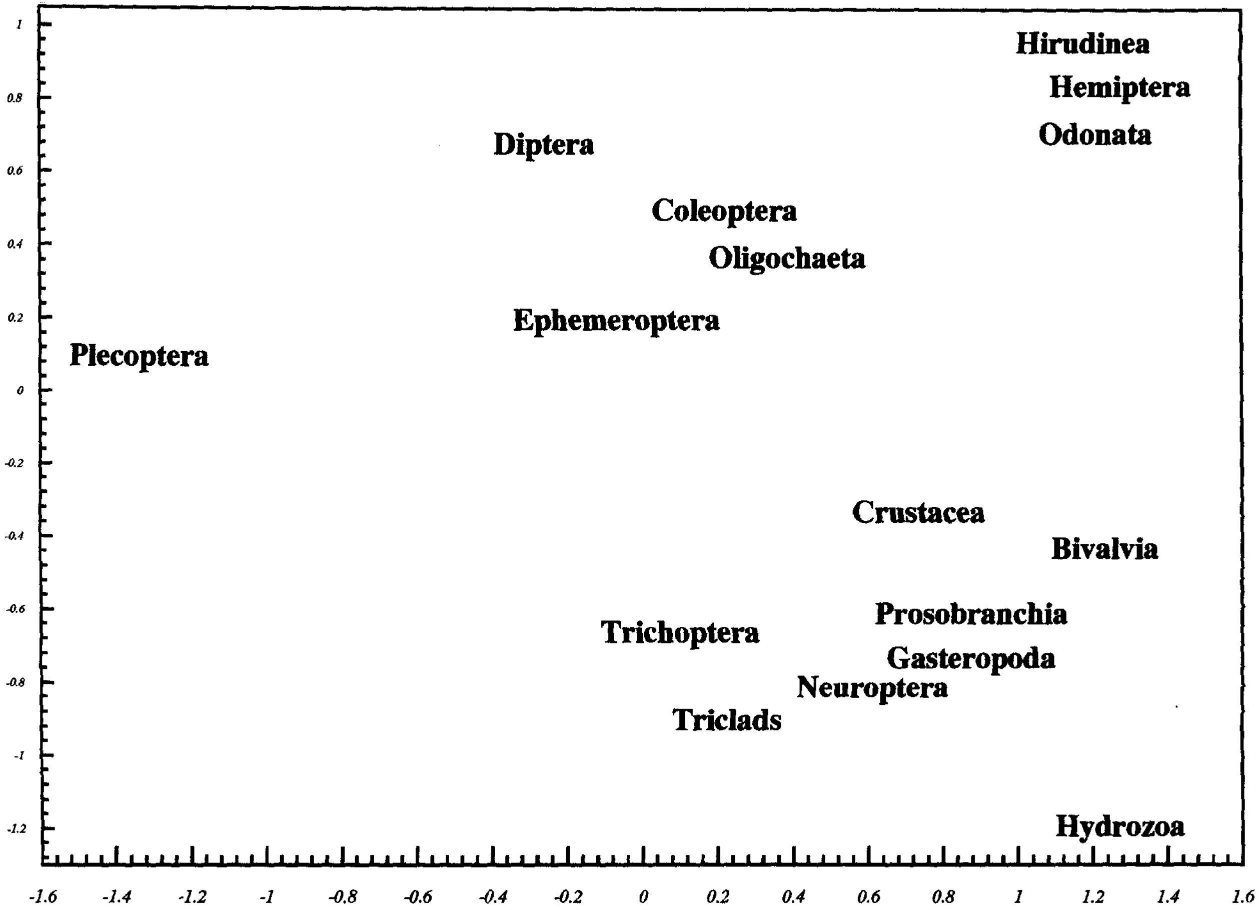

Fig. 3. diagram of taxa scores: abscissa is the first CA ordination axis, ordinate is the second CA ordination axis. Taxa groups labels are plotted in a position corresponding to the mean score value of taxa included in the group. For example mean scores of the genera of Plecoptera establish the position of the label 'Plecoptera'.

highly significant ($F \approx 13$ with 19 and 631 D.F.). This means that biotic indices considered are significantly related with the environmental variables included in the model. The predictors with the highest loadings are: particle size, slope, water conductivity and visual index (Table 5).

Discussion

Correspondence analysis results show that the first ordination axis is an upstream – downstream gradient. This is in agreement with other studies (Verneaux 1973).

The second axis separates taxa living in permanent waters (Triclads, Mollusca and Crustacea) from taxa living in temporary waters, mainly insects (Odonata, Hemiptera and Diptera). This is in agreement with the opinion that Peracaridans, Molluscs and Triclads dominate the hard water limestone springs and more generally in hard permanent waters, whereas insects dominate waters subject to flood and drying (Glazier, 1991).

Among insects only Trichoptera have a negative loading in the second axis. This can be interpreted as a preference of Trichoptera for permanent waters.

The importance of the upstream-downstream gradient and of temporary-permanent waters gradient was also observed by Verdonschot (*op cit.*) using DCCA.

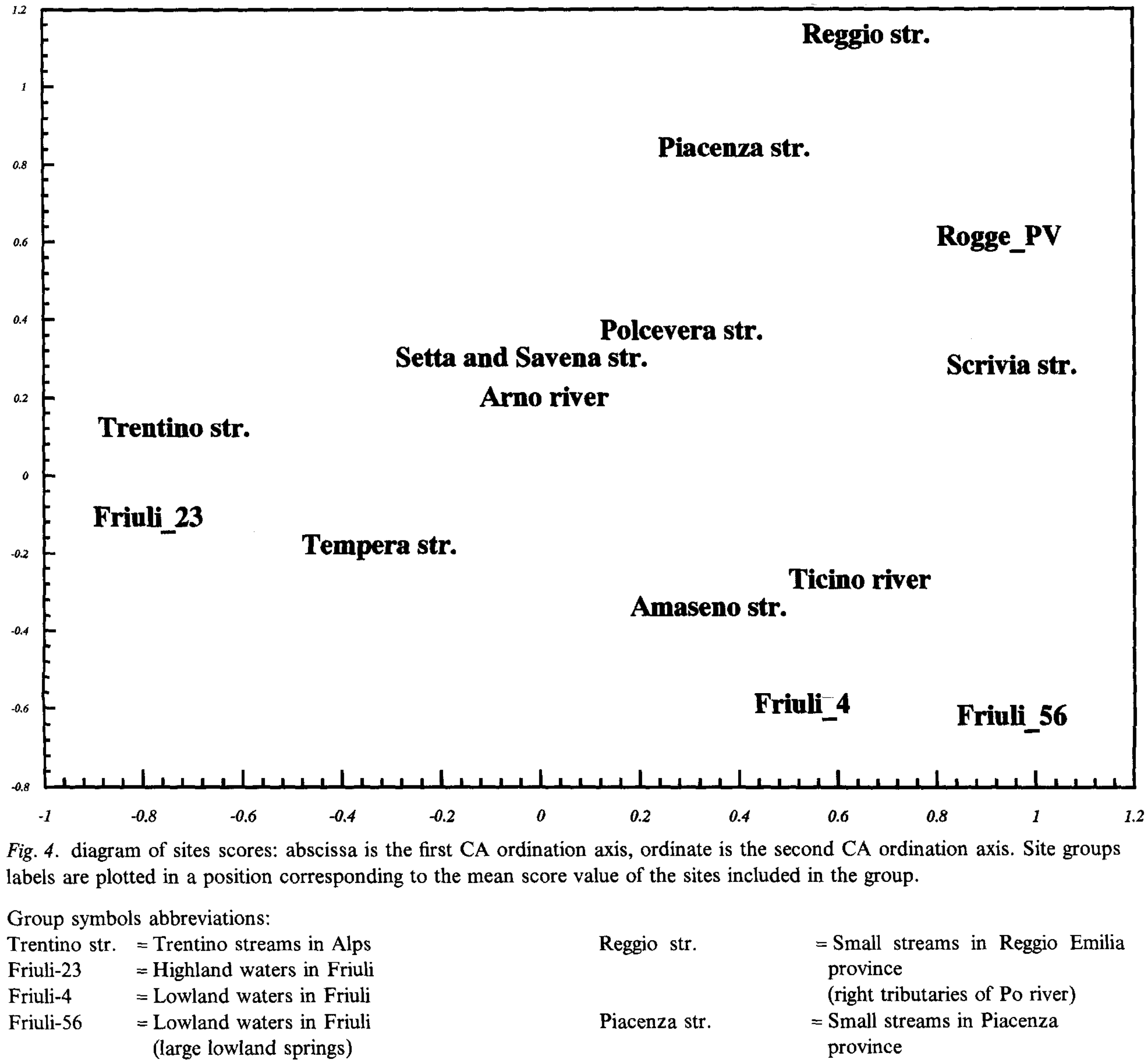

Fig. 4. diagram of sites scores: abscissa is the first CA ordination axis, ordinate is the second CA ordination axis. Site groups labels are plotted in a position corresponding to the mean score value of the sites included in the group.

Group symbols abbreviations:
Trentino str. = Trentino streams in Alps
Friuli-23 = Highland waters in Friuli
Friuli-4 = Lowland waters in Friuli
Friuli-56 = Lowland waters in Friuli (large lowland springs)
Amaseno str. = Amaseno and Vera streams in central Italy
Arno = Arno river in Toscana
Ticino river = Ticino river (left tributary of Po river)
Rogge-PV = Artifical channels near Pavia (Ticino river)
Reggio str. = Small streams in Reggio Emilia province (right tributaries of Po river)
Piacenza str. = Small streams in Piacenza province (right tributaries of Po river)
Scrivia str. = Scrivia stream
Polcevera str. = Polcevera stream
Setta and Savena str. = Streams near Bologna (Apennines)

Biotic indexes are well correlated with ordination axes (Table 4), suggesting that both biotic indices and ordination axes are deeply influenced by natural physical factors, so they cannot be considered '*per se*' a valid measure of water quality.

The ability of biotic indices to summarize water quality has been the object of some debate in the past (Armitage *et al.*, 1983). This brought many countries to develop new indices (Armitage *et al.*, 1987; Moss *et al.*, 1987; Metcalfe, 1989). For example low values of biotic indices were observed in waters with a low nutrient content, for this reason Stoch (1986) developed a new index (IRE)

Table 5. Multiple regression analysis results.

	ASPT	EBI
DETERMINANT	$0.52465 * 10^{-2}$	$0.52465 * 10^{-2}$
MULTIPLE R SQUARE	0.283	0.280
MULTIPLE R	0.532	0.529
F FOR ANALYSIS OF VARIANCE ON R	13.101	12.894
N.D.F.1	19	19
N.D.F.2	631	631
INTERCEPT CONSTANT	4.394	5.983

	ASTP		EBI	
	β	Structure R	β	Structure R
Particle size	0.19	0.56	0.09	0.44
Part. size heterog.	- 0.01	0.07	0.04	- 0.01
Slope	0.13	0.41	0.07	0.19
Source distance	- 0.37	0.09	- 0.31	0.31
River catchment	0.17	0.31	0.11	0.37
River discharge	0.02	0.15	0.09	0.29
River disch. summer	0.06	0.27	0.08	0.33
River disch. at outlet	0.16	0.28	0.01	0.17
Current speed	0.03	0.12	0.05	0.09
Total length	- 0.09	0.27	0.13	0.45
Wetted stream wide	0.43	0.14	0.38	0.37
Dry stream wide	- 0.07	- 0.12	0.02	0.06
Depth	- 0.03	- 0.11	0.07	0.09
Water temp.	- 0.11	- 0.24	- 0.10	- 0.16
Canopy cover	- 0.01	- 0.03	0.08	0.13
EC	- 0.24	- 0.47	- 0.26	- 0.48
BOD	- 0.10	- 0.22	- 0.12	- 0.25
Equivalent habit.	0.01	0.08	- 0.09	- 0.10
Visual index	0.16	0.48	0.17	0.47

to offset the poor performance of EBI when applied to Friuli waters. Samples from soft substrates often give also a low value of EBI. This is confirmed by the present data analysis.

Multiple regression confirm that biotic indices (Table 5) are highly related with physical environmental variables.

Armitage *et al.* (1983) suggested to provide target values of biotic indices (BMWP and ASPT) using multiple regression with physical and/or chemical variables against which observed values can be compared.

The advantage of the use of CA is that taxa and site scores can be easily computed and no external information is required to obtain indicator values of taxa. Computations can be done with presence-absence data, even if abundance of taxa aids in strength of the analysis.

The limit of the use of CA site scores instead of biotic indices is that scores can be calculated only as relative values within a determined data set, so target scores must be recalculated whenever new sites are included in the data set.

The level of taxonomic accuracy used in routine macroinvertebrate analysis (in the present analysis the determination of taxa was to the genus or to the family level) is enough to evaluate environmental quality only in routine work. Organic pollution can transform a bare gravel substrate into an organic rich soft bottom and this results in a change in genera and families present. When pollution is moderate, substitution of tolerant species with non tolerant ones can occur without changes at the genus or family level and both CA data analysis and biotic indices can fail to detect environmental change.

Future research need is testing both multivariate methods and biotic indices in different ecological conditions, areas and taxonomic detail.

Acknowledgements

This work was supported by a grant from Italian Ministero della Università e della Ricerca Scientifica e Tecnologica, (40% funds) and C.N.R..

References

Armitage, P. D., D. Moss, J. F. Wright & M. T. Furse, 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running water sites. Wat. Res. 17: 333–347.

Armitage, P. D., R. J. M. Gunn, M. T. Furse, J. F. Wright & D. Moss, 1987. The use of prediction to assess macroinvertebrate response to river regulation. Hydrobiologia 144: 25–32.

Cooley, W. W. & P. R. Lohnes, 1973. Multivariate data analysis, John Wiley, New York.

Crunkilton, R. L. & R. M. Duchrow, 1991. Use of stream order and biological indices to assess water quality in the Osage and Black river basins of Missouri. Hydrobiologia 224: 155–166.

De Pauw, N. & G. Vanhooren, 1983. Method for biological water quality assessment of water courses in Belgium. Hydrobiologia 100: 153–168.

Glazier, D. S., 1991. The fauna of North American temperate cold springs: patterns and hypotheses. Freshwat. Biol. 24: 527–542.

Ghetti, P. F., 1986. I macroinvertebrati nell'analisi di qualità dei corsi d'acqua. Provincia Autonoma di Trento. ISSN 0393-5736: 111 pp.

Hilsenhoff, W. L., 1988. Seasonal correction factors for the biotic index. The great lakes entomologist 21: 9–13.

Metcalfe, J. L., 1989. Biological water quality assessment of running waters based on macroinvertebrates communities: history and present status in Europe. Envir. Poll. 60: 101–139.

Miller, M. C. & J. R. Stout, 1989. Variability macroinvertebrate community composition in an arctic and subarctic stream. Hydrobiologia 172: 111–127.

Minshall, W. G. & R. C. Petersen, 1985. Towards a theory of macroinvertebrates community structure in a stream ecosystem. Arch. Hydrobiol. 104: 49–76.

Moss, D., M. T. Furse, J. F. Wright & P. D. Armitage, 1987. The prediction of the macro-invertebrate fauna of unpolluted running water sites in Great Britain using environmental data. Freshwat. Biol. 17: 41–52.

Rohlf, F. J., 1985. NTSYS, numerical taxonomy system of multivariate statistical programs. The State University of New York at Stony Brook.

Sheldon, A. L. & R. A. Haick, 1981. Habitat selection and association of stream insects: a multivariate analysis. Freshwat. Biol. 11: 395–403.

Stoch, F., 1986. Nota preliminare su una nuova metodologia biologica per il mappaggio di qualità delle acque correnti, Acqua-Aria 2: 137–142.

Ter Braak, C. J. F. & I. C. Prentice, 1988. A theory of gradient analysis. Adv. Ecol. Res. 18: 271–315.

Ter Braak, C. J. F., 1990. Update notes: CANOCO Version 3.10, Agriculture Mathematics Group. Wageningen: 35 pp.

Verdonschot, P. F. M., 1990. Ecological characterization of surface waters in the province of Overijssel (The Netherlands). Rep. Res. Inst. Nature Management Leersum, 1990: 255 pp.

Verneaux, J., 1973. Recerches écologiques sur le réseau Hydrographique du Doubs – essai de biotypologie. Ministère de l'Agriculture: 261 pp.

Wartenberg, D., S. Ferson & F. J. Rohlf, 1987. Putting things in order: a critique of detrended correspondence analysis. Am. Nat. 129: 434–448.