Aquatic Ecology (2007) 41:387–398
DOI: 10.1007/s10452-005-2836-0

Application of clustering techniques for the characterization of macroinvertebrate communities to support river restoration management

V. Adriaenssens[1,*], P.F.M. Verdonschot[2], P.L.M. Goethals[1] and N. De Pauw[1]
[1]*Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, 9000 Gent, Belgium;* [2]*Alterra Green World Research, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands; *Author for correspondence (e-mail: Peter.Goethals@UGent.be)*

Received 30 March 2004; accepted in revised form 24 February 2005

Key words: European Water Framework Directive, Multivariate analysis, Ordination, River typology, Self-organizing maps, Similarity ratio

Abstract

The European Water Framework Directive prescribes that the development of a river assessment system should be based on an ecological typology taking the biological reference conditions of each river type as a starting point. Aside from this assessment, water managers responsible for river restoration actions also need to know the steering environmental factors to meet these reference conditions for biological communities in each ecological river type. As such, an ecological typology based on biological communities is a necessity for efficient river management. In this study, different clustering techniques including the Sørensen similarity ratio, ordination analysis and self-organizing maps were applied to come to an ecological classification of a river. For this purpose, a series of sites within the Zwalm river basin (Flanders, Belgium) were monitored. These river sites were then characterized in terms of biotic (macroinvertebrates), physical–chemical and habitat variables. The cluster analysis resulted in a series of characteristic biotic communities that are found under certain environmental conditions, natural as well as human-influenced. The use of multiple clustering techniques can be of advantage to draw more straightforward and robust conclusions with regard to the ecological classification of river sites. The application of the clustering techniques on the Zwalm river basin, allowed for distinguishing five mutually isolated clusters, characterized by their natural typology and their pollution status. On the basis of this study, one may conclude that river management could benefit from the use of clustering methods for the interpretation of large quantities of data. Furthermore, the clustering results might enable the development of a cenotypology useful for efficiently steering river restoration and enabling river managers to meet a good ecological status in most of the rivers as set by the European Water Framework Directive.

Introduction

The ecological classification and delineation of river communities is a tool that serves as a basis for river assessment and management. By knowing what should be the original biological community at a river site, one can assess the degree to which human activities have altered it (Hawkins et al. 2000). This definition of the reference conditions by means of biological communities is essential to set up a biological river assessment system. Beside this ecological evaluation objective, river restoration endpoints are also often defined within a community ecology perspective (Palmer et al. 1997).

More recently, ecological classifications serve as a base for river management imposed to the EU member states by the Water Framework Directive (EU 2000). The Directive distinguishes two key goals for rivers: (1) programmes of measures to achieve at least a 'good ecological status', and (2) a management system based upon natural river basin districts. To reach these goals, EU member states need to implement an assessment system with type-specific ecological reference conditions as well as a type-specific river management. Therefore, within each river basin district, all water bodies must be classified according to an ecologically relevant typology (Chave 2001). For the biotic assessment of rivers in Flanders (Belgium), the Belgian Biotic Index (BBI) (De Pauw and Vanhooren 1983; De Pauw and Vannevel 1991) and the Belgian Sediment Index (BSI) (De Pauw and Heylen 2001), which both are based on diversity and tolerance of macroinvertebrate taxa, are currently used as a standard tool. However, these indices are not yet differentiated per river type but applied in a uniform way in Flanders. River management on the other hand is based on legislative hydrological boundaries and organized within river basins (Schneiders and Verheyen 1998). This hierarchical organization of river systems by the water boards in Flanders, is based on merely physical and socio-political perspectives. As a result it neglects scales relevant to the biota, while biological information should be included as a primary hierarchical component (Parsons et al. 2003).

In community ecology, one has been debating whether or not communities can be described as units that are discrete, clearly defined, and integrative (i.e. defined by interactions). In many cases, community boundaries are more or less arbitrarily set by ecologists within the scope of their study and experience. In general, the level of ecological organization and the aspect of community definition are very scale-dependent (Palmer and White 1994; Palmer et al. 1997). To classify sites into clusters, we may apply fuzzy (ordination) analysis or crisp classification (sensu stricto clustering). Fuzzy classifications allow observing the species gradients that may exist between communities.

In this study, three techniques have been applied which differ in their approaches regarding clustering algorithms, their theoretical basis with regard to the assumed biotic type of response to environmental conditions and their potential value as a tool for river management. The applied agglomerative clustering method based on the Sørensen similarity ratio (SR) (Van Tongeren 1986) provides a crisp classification as given by the majority of clustering techniques (Cao et al. 1997; Halkidi et al. 2001). Ordination (Hill 1979) on the other hand is based on a gradient analysis. This multivariate technique allows for detecting sample groups with a similar species composition and relating observed patterns with environmental variables (Pardo and Armitage 1997). A third clustering technique, the self-organizing maps (SOMs), is since recently becoming popular in ecology because of its advantage to deal with non-linear and heterogeneous data (Foody 1999; Walley et al. 2000).

The main objective of the present study was to compare the outcome of different clustering techniques and to link the resulting ecological classification to river types and their characterizing environmental conditions, natural as well as human-impacted. For this study, a dense monitoring network in a river basin in Flanders was used. The classification was based on a scale relevant to macroinvertebrate communities. Macroinvertebrates were used to construct this typology because they are known to be good indicators of the ecological quality of a river and they can integrate changes in environmental conditions over time (Hawkes 1979; Rosenberg and Resh 1993).

Material and methods

Study area and data collection

The study area was the Zwalm river basin (Figure 1) which is part of the Upper-Scheldt, covering a total surface of 11.650 ha. The Zwalm river itself has a length of 22 km. The southern part of the Zwalm basin consists of small brooks located in the crenal zone, where groundwater flows in the brooks at the source. These brooks are expected to be unpolluted. Because of the specific geomorphology in this area, they are characterized by a unique fauna (Goethals and De Pauw 2001). On the other hand, major parts of the river basin are

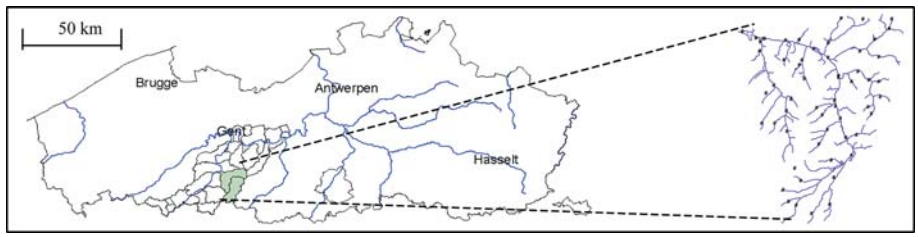

Figure 1. The Zwalm river basin, located in the Upper-Scheldt basin in Flanders (Belgium).

Table 1. Environmental variables measured in the Zwalm river basin.

Variables	Measuring units	Method
Temperature	°C	Oximeter WTW Oxi 330
pH	$-\log [H+]$	pH meter Consort P114
Conductivity	μS/cm	WTW 249 electrode
Suspended solids	mg/l	Filtration
Dissolved oxygen	mg/l	Oximeter WTW Oxi 330
Water level	cm	Field measurement
Fraction pebbles	% surface bottom	Van Veen grab/sieving method
Fraction sand	% surface bottom	Van Veen grab/sieving method
Shadow	% surface bottom	Field measurement/visual
Aquatic macrophytes	Presence/absence	Field measurement/visual
Width	cm	Field measurement
Stream velocity	m/s	Field measurement/hydrometric propeller
Embankment	2 categories (0 (absent), 1 (partial), 2 (total))	Visual observation according to Schneiders et al. (1999)
Meandering	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders et al. (1999)
Hollow banks	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders et al. (1999)
Pools/riffles	6 categories (1 (well developed) to 6 (absent))	Visual observation according to Schneiders et al. (1999)

significantly impacted mainly by untreated urban wastewater and diffuse pollution originating from agricultural activities (Goethals and De Pauw 2001). Habitat degradation of the watercourses in this region is mainly caused by artificial embankments, weirs and erosion.

Sixty sites within the Zwalm river basin were sampled in autumn 2000 and 2001. The site selection was based on a maximal variation in natural and anthropogenic characteristics. Biological sampling consisted of collecting macroinvertebrates using a standard handnet during 5 min kick-sampling over a river stretch of 10 m (IBN 1984), thereby exploring in a representative way the different habitats present (De Pauw and Vanhooren 1983). This approach can also be described as a multihabitat sampling (Barbour et al. 1999). In addition, a number of environmental variables and habitat characteristics were measured (Table 1). Macroinvertebrate samples were identified up to group, family or genus level as defined by De Pauw and Vanhooren (1983).

Data analysis

Data analysis was first performed by means of SR clustering and SOMs. These clustering results were then plotted along the first two axes of an ordination diagram based on the results of an indirect ordination analysis (Detrended Correspondence Analysis). As a second step, the characterizing environmental conditions of each cluster were taken into consideration by means of a direct ordination analysis (Detrended Canonical Correspondence Analysis) and SOMs.

Before data analysis, the abundance values of the total number of macroinvertebrates per sample and values of environmental variables (except pH and the categorical variables like meandering, pool/riffle and hollow bed development) were log $(x + 1)$ transformed.

In the following paragraphs, detailed information is given on the applied clustering techniques.

SR clustering with the programme FLEXCLUS (Van Tongeren 1986) is based on the Sørensen

similarity ratio (Sørensen 1948) (formula 1).

$$SR_{ij} = \sum_k y_{ki} y_{kj} / \times \left(\sum_k y_{ki^2} + \sum_k y_{kj^2} - \sum_k y_{ki} y_{kj} \right) \quad \text{formula 1}$$

y_{ki} = the abundance of the kth species at site i and y_{kj} = the abundance of the kth species at site j.

The initial step in FLEXCLUS is a non-hierarchical clustering that handles noise and redundancy by combining samples into groups following the algorithm of Sørensen (1948). Samples are fused into clusters when their similarity is higher than a given threshold value. Refinement of the initial clustering by reallocation leads to a final clustering, which is a combination of fusion and division of clusters based on the distance of a sample to the cluster centroid. The ordering of clusters is obtained by reciprocal averaging (Hill 1973; Van Tongeren 1986).

In this study, the objective was to extract clusters characterized by a high degree of isolation. Therefore, for the following clustering parameters, the most optimal values were chosen: (1) the threshold for initial clustering (dissimilarity limit of the clustering), (2) the relocation of initial clustering and (3) the option down-weighting of rare taxa. The procedure used to achieve optimal clustering involved: (1) maximize the internal homogeneity (a measure of the similarity of samples within a cluster), (2) minimize the resemblance (the mean resemblance to another cluster), (3) maximize the degree of isolation (the dissimilarity of one cluster with the closest one), (4) maximize the stability (the amount of sample relocations during different clustering steps) and (5) minimize the number of clusters that contain only one sample.

Ordination was performed by means of the CANOCO programme version 4.02 (Ter Braak and Smilauer 1998). Detrended Correspondence Analysis (DCA) (Hill 1979) was used to perform an indirect gradient analysis of the Zwalm river basin sites. The percentage of cumulative variance of the species data was used as a parameter of the explanatory power of the canonical axes. The clusters resulting from the SR clustering were visualized in the DCA ordination diagram. The following DCA options were applied: detrending by segments and no down-weighting of rare taxa.

Detrended Canonical Correspondence Analysis (DCCA) was used to perform a direct unimodal gradient analysis by means of the CANOCO programme (Ter Braak and Smilauer 1998). In this analysis, habitat characteristics, such as, meandering, pool/riffle, and hollow beds, were included as categorical variables. The following DCCA analysis options were applied: no down-weighting of rare taxa and detrending by second-order polynomials.

Self-organizing maps (SOMs) (Kohonen 1982), programmed within MATLAB 5.3 by Vesanto et al. (2000), were used to cluster the dataset. SOM is an unsupervised neural network that identifies patterns in data, groups them into a predefined number of classes, and orders the classes in a two-dimensional output space such that near neighbours in a data space are near neighbours in an output space. Clustering was based on the Euclidean distance between different samples and a neighbourhood function ensuring that near neighbours in the output space represent similar patterns (cf. Walley and O'Connor 2001). First, the number of map units (3×3) and the size of the map was determined. The two highest eigenvalues of the data set were calculated and the ratio between side-lengths of the map grid was set to this ratio. Next, the actual side-lengths were set so that their product is as close to the desired number of map units as possible. Then the SOM was initialized. Linear initialization along the two highest eigenvectors was done. After initialization, the SOM was trained in two phases: first a rough training and then a fine-tuning. The training was done with the sequential training algorithm. The SOM was trained iteratively. In each training step, one sample vector x from the input data set was chosen randomly and the distances between it and all the weight vectors of the SOM were calculated using the Euclidian distance measure. The neuron whose weight vector was closest to the input vector x was called the Best-Matching Unit, denoted here by $c : \| x - m_c \| = \min_i \{ \| x - m_i \| \}$ where $\| \ \|$ is the Euclidean distance measure.

Results

Sixty six macroinvertebrate taxa were sampled during 2000 and 2001 in the Zwalm river basin and identified upon genus, group or family level as

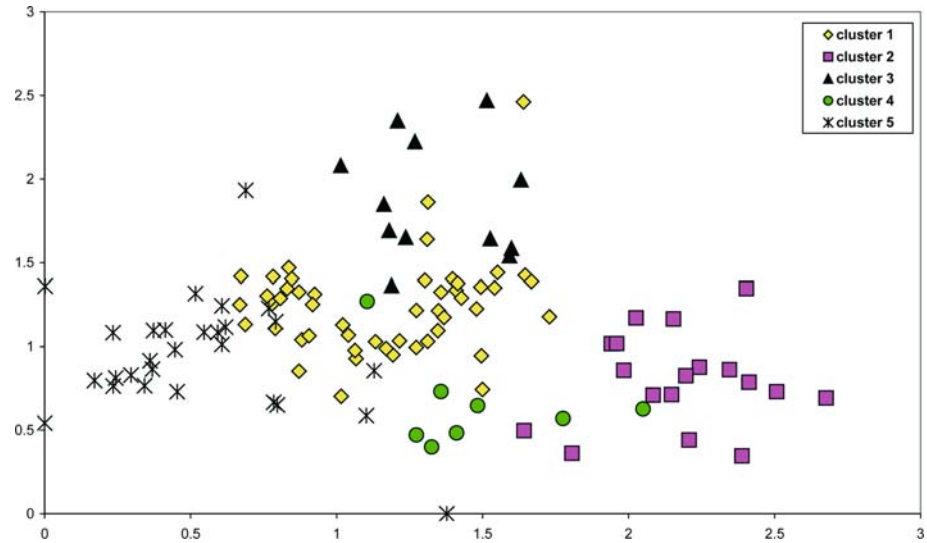

Figure 2. DCA plot of sites in the Zwalm river basin with their FLEXCLUS cluster membership indicated.

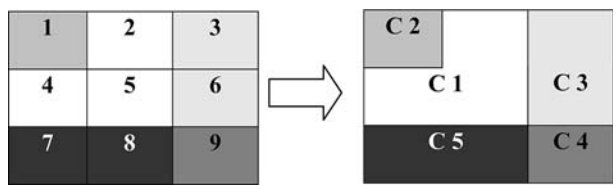

Figure 3. SOM rectangular grid with nine neurons each containing a set of river sites in the Zwalm river basin and grouped in a next stage into five clusters (C).

defined by De Pauw and Vannevel (1991). The SR clustering resulted in five clusters.

Sites within these clusters were visualized by means of plotting each site defined by its SR cluster membership at a position along the first and second axis of the DCA. In this indirect analysis (DCA), the percentage of variance explained by the first two canonical axes was relatively low, 9.3% and 5.9%, respectively. Examinations of the positions of the sites along the first two DCA ordination axes (Figure 2) revealed nearly the same clustering structure in the macroinvertebrate data as with the SR clustering.

The result of the SOM clustering is a map where each neuron represents values for each variable and adjacent neurons are characterized by more similar patterns than others. Each neuron can then be seen as the centroid of a river type, and the map as a classification of rivers based on the macroinvertebrate abundance data. In this way, a SOM can be seen as a combination of classification and multi-dimensional ordination. In this study, a SOM of nine neurons (3×3 rectangular grid) was created, by means of a non-linear projection of the data set onto a grid, in this case a rectangular grid (Figure 3). Finally, the nine neurons could be grouped into five clusters based on the preferences of sites for nearby neurons when repeating the SOM process ten times. Similar cluster groups were obtained with the SOM technique and the SR clustering.

Figure 4 shows the membership of sites in the Zwalm river basin with the cluster groups obtained by the SOM clustering, plotted along the first and second DCA axis.

To link the clustering results to the environmental variables, the biological data were analyzed together with the environmental data of the ordination analysis as well as those of the SOMs. Indirect ordination by means of DCA gave a length of gradient of the first axis of 2.677. As the eigenvalues of the first and the second axis of the

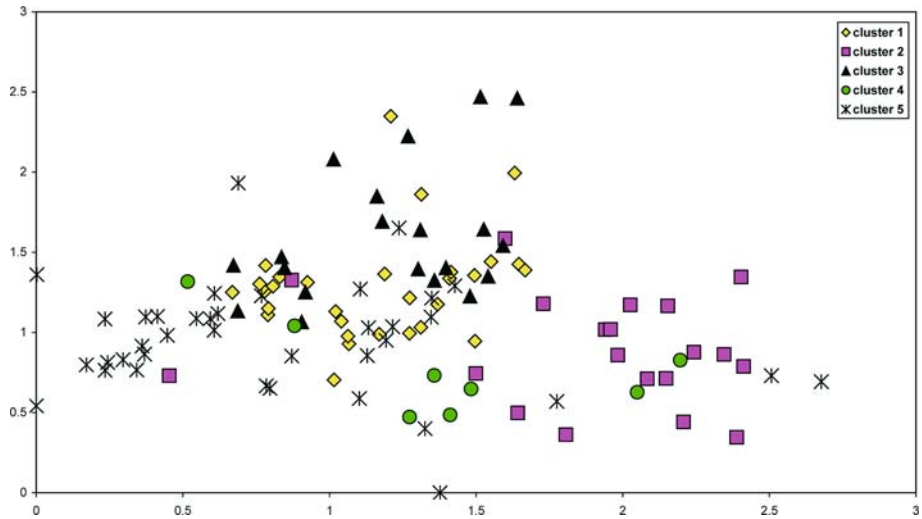

Figure 4. DCA plot of sites in the Zwalm river basin with the SOMs cluster membership indicated.

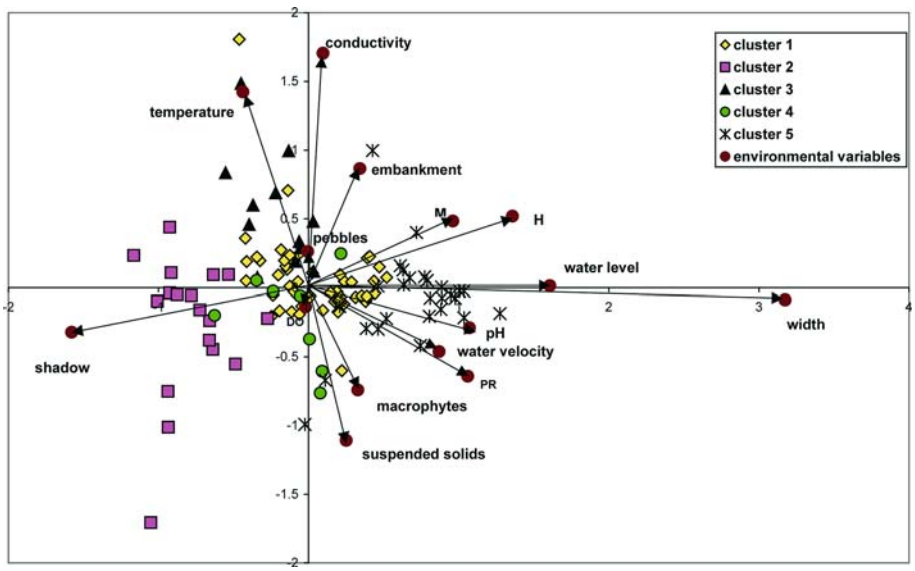

Figure 5. DCCA biplot of samples with environmental variables (DO = dissolved oxygen concentration, PR = pool/riffle development; M = meandering development, H = hollow bank development), grouped by their SR cluster membership.

DCCA analysis were only slightly lower than these of the DCA analysis, the measured environmental variables explain a large part of the variation of the data. The percentage of variance of species-environment relations in DCCA were 29.4 and 14.6% for axis 1 and 2, respectively. These high percentages indicate the explanatory strength of the environmental variables. The variables width, meandering and conductivity explain the major part of the variance of the macroinvertebrate data, as can be seen in the biplot of samples (grouped by their SR cluster membership) and environmental variables with the 1st and 2nd DCCA axis (Figure 5).

The biplot (Figure 5) shows that width and conductivity explain an important part of the variance in the macroinvertebrate data. Moreover, sites plotted along the first axis are revealing a gradient from upstream to downstream, while the second axis is correlated with an increasing

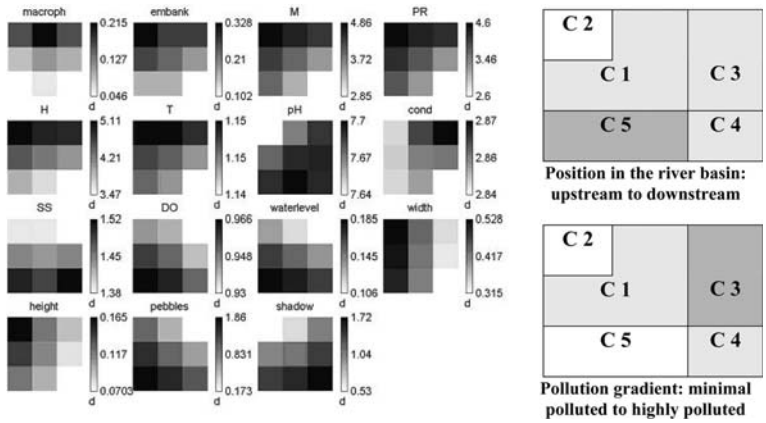

Figure 6. SOMs plot showing the five clusters with the nine neurons containing the different sites in the Zwalm river basin represented by their environmental variables (PR = pool/riffle development; M = meandering development, H = hollow bank development, T = temperature, cond = conductivity, SS = suspended solids, DO = dissolved oxygen concentration, C = Cluster).

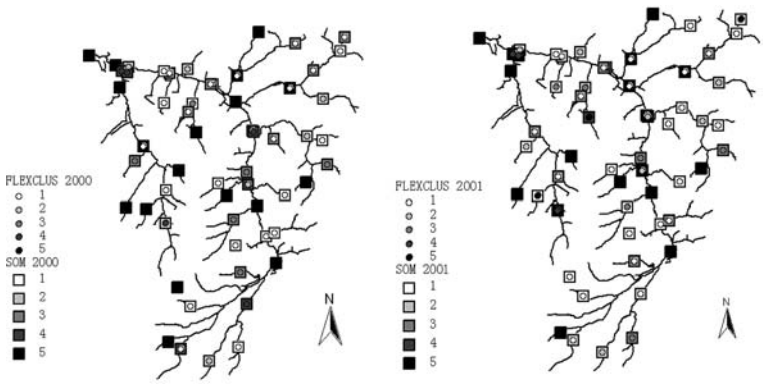

Figure 7. Overview of the membership of samples to the clusters for the different sampling years (left: 2000, right: 2001) as obtained by SR clustering and SOM clustering.

pollution pressure (e.g. conductivity) and habitat degradation (e.g. embankment). Samples from the same sampling site but a different year (2000/2001) were mostly grouped together in the clusters (Figure 7). Most sampling sites contained a high number of dominant and very general taxa e.g., Chironomidae *non-thummi plumosus*, Tubificidae, and Gammaridae. This suggests that clusters were separated from each other based on less frequently occurring taxa.

With regard to the SOM technique, the environmental gradient could be shown by means of a two-dimensional plot visualizing the measured environmental variables over the nine neurons (Figure 6).

In parallel with the ordination analysis, the main gradients explaining the variance in biological communities found by SOMs were the position in the river basin coupled to the dimension of the river and the pollution gradient (e.g. conductivity) (Figure 5).

Figure 7 shows that the membership of sites to the cluster groups obtained with the SOM clustering were for a major part similar to the obtained clusters with the SR clustering.

Based on the explanatory environmental variables and characterizing macroinvertebrate taxa given by the results of the three clustering techniques, several ecological river types could be distinguished and characterized, containing both natural as well as degraded river conditions (Table 2).

Based on the explanatory environmental variables and characterizing macroinvertebrate taxa

Table 2. Environmental characteristics of the obtained clusters based on the ecological classification of sites in the Zwalm river basin.

	Position in the river basin	Macroinvertebrate taxa	Physical–chemical variables	Habitat characteristics
1	Mainly upstream sites in the main headstream	Gammaridae Tubificidae Chironomidae *thummi plumosus* Chironomidae *non-thummi plumosus*	Not characterized by any specific environmental variable	
2	Brooks in the crenal zone and small brooks	High taxon diversity Trichoptera Ephemeroptera Tubificidae Chironomidae *non-thummi plumosus* Gammaridae	Low conductivity, shadowy, high concentrations of suspended solids	High habitat variability
3	Mainly small brooks in agricultural areas	*Gyraulus* *Armiger* Chironomidae *non-thummi plumosus* Tubificidae	High conductivity	Indifferent
4	Disperse distribution of sites	*Nepa* *Gyraulus* Asellidae Gammaridae Hydropsychidae Chironomidae *Erpobdella* Tubificidae	Low conductivity	Stony surface, high stream velocity, shadowy
5	Mainly downstream sites in the main headstream	Asellidae Chironomidae *thummi plumosus* Chironomidae *non-thummi plumosus*	Heterogeneous cluster for most of the physical–chemical variables	High water level and width

given by the results of the three clustering techniques, several ecological river types could be distinguished and characterized, containing both natural as well as degraded river conditions (Table 2).

Discussion

There have been numerous debates on whether or not communities can be described as units that are discrete and clearly defined (Palmer et al. 1997). Is a community 'set' by ecologists in order to study operationally this level of ecological characterization or is this a real level of organization? This discussion is also embedded in the community-continuum debate (McIntosh 1967; Whittaker 1967; Austin 1985; Austin and Smith 1989; Collins et al. 1993). This means that a species response to environmental variables is a gradient response, and the delineation of a community is time and space dependent. As such, a community is not a strict entity, but an entity that has fuzzy boundaries over the different environmental variables with time and space as extra dimensions. Because stream systems are often viewed as harsh environments containing fewer competitive dominants and therefore having less need for interspecific resource partitioning (Townsend 1989; Hildrew 1992), environmental influences are more important than biotic interactions structuring biotic communities. As such, the authors define a macroinvertebrate community as a group of macroinvertebrate taxa that are able to maintain a population under the set specific environmental conditions. The community can serve as an entity used by river managers to describe the biological state of a river at certain environmental conditions. Clustering techniques can be useful for river managers to delineate communities that can serve as an object of river site characterization, assessment and restoration, taking into account the fuzzy boundaries of the community itself.

The results of SR, DCA and SOM clustering show that clusters of sites in the Zwalm river basin cannot easily be distinguished on the basis of the available data of macroinvertebrate communities. The most important causes are the strong influence of diffuse agricultural pollution in a high number of the sampling sites in the Zwalm river basin and

the relative high degradation of the habitat characteristics of the watercourses. This is shown by the high values of conductivity and suspended solids and the low structural diversity. Key taxa that define the clusters were very common and abundant taxa like Gammaridae, Asellidae, Chironomidae and Tubificidae (Table 2).

The major explanatory environmental variables of the macroinvertebrate communities in the studied streams are dimension-related, reflected in environmental variables such as width, flow velocity and water level, and distance to mouth. As stream size increases throughout the Zwalm river basin, taxa that are specific for wider and deeper streams dominate (Table 2) (Verdonschot 1995), in correspondence to the River Continuum Concept (Vannote et al. 1980). A second important variable found in this analysis can be related to human impact caused by agricultural activities, as is reflected by an increase in conductivity. Although river enrichment should not only be measured by means of conductivity, this study shows the relevance of conductivity over dissolved oxygen concentration measurements. As was illustrated in Figure 5, the dissolved oxygen concentration does not contain a large explanatory capacity with regard to the distribution of biological communities in the Zwalm river basin, and this as a result of the relative high flow velocities and related mixing. However, in Flanders, dissolved oxygen is still used as the main variable for physical–chemical water quality assessment which is expressed as the Prati Index for dissolved Oxygen (PIO) (cf. Prati et al. 1971; VMM 1997). Discrete measurements of dissolved oxygen to calculate these PIO indices are consequently insufficient to draw conclusions concerning their effect on biological life in the Zwalm river basin, while conductivity measurements seem to be less variable in time and giving a good indication of the pollution load (Vandenberghe et al. 2004).

To set up a stream classification, often only one approach or technique is used to analyze community data. This can however be rather subjective (Jackson 1993). For that reason, in our study, different clustering techniques were used. These techniques were assessed based on different criteria: (1) the interpretability of the clustering results, (2) the way of interpretation of the ecological communities, (3) the subjectivity involved in the preferred parameter settings and interpretation of the results, (4) the theoretical assumption underlying the biotic response, and (5) the usefulness in river management as a decision support tool.

The first criterion, considering the interpretability of the clustering results, showed that SR clustering revealed nearly the same structures as the ordination analysis (Figure 3), but defined groups by means of the SOM showed more overlap in the ordination diagram (Figure 5). Smoothening between the clusters is apparent in the results of each clustering approach that is mainly caused by river pollution and habitat degradation and the absence of natural reference conditions for most of the river types. However, all cluster techniques resulted in a relative interpretable cluster structure, including natural river types as well as river types characterized by specific environmental conditions (Table 2).

Based on the second criterion, looking at the representation of ecological communities it was clear that gradient analysis as a 'fuzzy' classification method was less subjective in contrast to the 'crisp' classification by SR and SOMs. Moreover, within the ordination analysis, the ecological relevance of this discovered gradient is given by the calculation of the explained variation in the data set based on the biological data, which can also be seen as an advantage considering the third criterion, evaluating the subjectivity involved. Subjectivity is apparent in the SR clustering when a lot of *a priori* decisions have to be made in advance during the clustering process as described by Van Tongeren (1986). This can be seen as an advantage as experts can intervene in the clustering process, but could become negative when 'subjective' pre-assumptions determine the cluster structure. Although the advantage of the lack of pre-assumptions in the SOM technique is often stressed (Brosse et al. 2001), the combination of the neurons into clusters in this study was however based on a relative subjective basis. Although only a few techniques have been proposed to detect cluster boundaries in a SOM, it is still recognized as a difficulty (Giraudel and Lek 2003). However, the unified matrix (U-matrix) approach (Ultsch and Siemon 1990) offers a way to detect a clustering by computing the distance between the sites within different neurons. Also high value distances can be used here as an indication of cluster boundaries. This technique might offer a more objective method for cluster analysis by SOMs in the future.

A sensitivity analysis of the SOM and U-matrix can then be computed to test the significance of each environmental variable on SOM results.

Based on the fourth criterion regarding the theoretical assumption of biotic response, the three used clustering techniques clearly differed in an important way. Because habitat preferences of taxa are often non-linear functions of habitat variables, linear techniques are therefore mentioned as not appropriate (Ter Braak and Verdonschot 1995). As such, it is suggested that important gradients explaining the macroinvertebrate distribution can be extracted by means of analysis of non-linear gradients, such as multivariate analysis with the unimodal (D)(C)CA ordination and non-linear SOMs, in contrast to the SR clustering.

Within the last criterion, evaluating their potential use as a decision support tool in river management, the visual representation of the different SOM clusters explains for a large part the recent success of these techniques to explain huge and variable data sets in ecology, in contrast to the more classical multivariate techniques as ordination analysis.

In this study, part of the sites was given a different cluster membership when comparing SR and SOMs clustering results. These sites, when plotted in the ordination diagram, were located in the overlap zone between the different clusters defined by both techniques (Figure 3, Figure 5). Ordination analysis allowed thus to interpret these clustering results on a more objective way with regard to the smoothening of clusters as a result of the present pollution gradient and the resulting uncertainty of classification. This information would be lost when only relying on a crisp classification technique such as SR and SOMs. The use of multiple clustering techniques and particularly the combination of gradient analysis with a crisp clustering method can thus be of advantage to draw more reliable and robust conclusions with regard to the ecological classification of river sites.

For an objective evaluation however, techniques need to be used to test statistically the difference between the different clustering methods. One suggestion for this is to use MANOVA (Multivariate Analysis of Variance), a technique that can compare samples based on two dependent variables (Cooley and Lohnes 1971).

When considering the evaluation of these clustering techniques, one has to take into account the rather limited deduction capacity of the used dataset. First, seasonal effects have not been considered although a dataset containing both spring and autumn samples is seen as optimal (Furse et al. 1984; Ruse 1996; Verdonschot 2000). The taxonomic resolution at a species-level classification is desired for an in-depth analysis of macroinvertebrate responses to environmental variables (Furse et al. 1984; Pardo and Armitage 1997; Verdonschot 2000; Adriaenssens et al. 2004). Also, in this study, a relatively small data set was used and as such the results may not be applicable on data sets covering whole Flanders. Furthermore, natural 'reference' brooks were nearly absent and the typology approach only concerned macroinvertebrates.

With regard to the integration of natural as well as degraded river conditions into a typology Verdonschot (1990) has been developing a so called cenotypology for use in river management in the Netherlands (Verdonschot 1995; Verdonschot and Nijboer 2000, Verdonschot et al. 2000; Verdonschot and Nijboer 2002). A cenotype is a site group that is established if it is clearly recognizable along an identified environmental gradient and if it has a distinct fauna. The objectives of management activities such as restoration can be set according to these cenotypes. By combining the biotic and abiotic parameters, a relation can be established between the desired biotic communities as final target values and the measures needed to achieve the target conditions. The directions of development from one type to the other are indicated by their (supposed) most important steering factor and will lead to restoration or mitigation actions. However, strong human influences on the water systems in Flanders have lead to the disappearance of communities that could serve as target or reference conditions. For reasons of efficient water management, reference conditions should therefore be known in terms of biological communities, and steering factors to obtain these conditions should be unravelled as well. A possibility could be to define these natural (reference) states by means of predictive modeling (Verdonschot 2000; Goethals and De Pauw 2001).

Evidently, the description of a clear cenotypology for running waters in Belgium and other EU countries will be an essential requirement to fulfil

the aims of the Water Framework Directive with regard to river assessment and river management.

Conclusions

An ecological classification of sites based on macroinvertebrate communities has been described for the Zwalm river basin in Flanders, Belgium. To this end, three clustering techniques were applied, one based on ordination, one on the Sørensen similarity ratio and one on self-organizing maps. Each of these techniques allowed for distinguishing five mutually isolated clusters of sites. The advantage of the use of multiple clustering techniques and particularly the combination of gradient analysis with a crisp clustering method enables making a more objective ecological classification of river sites. The obtained clusters, based on macroinvertebrate communities, could be linked with their natural typology and pollution status. The application of these clustering techniques on macroinvertebrate data might enable the establishment of a cenotypology which could be helpful to efficiently steer river restoration towards a good ecological status as set by the European Water Framework Directive.

Acknowledgements

V. Adriaenssens holds a grant from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). Part of the research has been carried out within the EU COST 626 Action (European Aquatic Modelling Network).

References

Adriaenssens V., Simons F., Goddeeris B., NguyenThi Hong L., Goethals P.L.M. and De Pauw N. 2004. Potential of bioindication of chironomid communities for assessment of running water quality (Flanders, Belgium). Belg. J. Zool. 134(1): 31–40.

Austin M.P. 1985. Continuum concept, ordination methods and niche theory. Annu. Rev. Ecol. Syst. 16: 39–61.

Austin M.P. and Smith T.M. 1989. A new model for the continuum concept. Vegetatio 83: 35–47.

Barbour M.T., Gerritsen J., Snyder B.D. and Stribling J.B. 1999. Rapid Bioassessment Protocols for use in Wadable Streams and Rivers: Periphyton, Benthic Macroinvertebrates and Fish, 2nd Ed. EPA 841-B-99-002. USEPA, Office of Water, Washington, DC.

Brosse S., Giraudel J.L. and Lek S. 2001. Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages. Ecol. Model. 146: 159–166.

Cao Y., Bark A.W. and Williams W.P. 1997. A comparison of clustering methods for river benthic community analysis. Hydrobiologia 347: 25–40.

Chave P. 2001. The EU Water Framework Directive, an introduction. IWA Publishing, 208p.

Collins S.L., Glenn S.M. and Roberts D.W. 1993. The hierarchical continuum concept. J. Veg. Sci. 4: 391–413.

Cooley W.W. and Lohnes P.R. 1971. Multivariate Data Analysis. John Wiley and Sons, Inc.

De Pauw N. and Vanhooren G. 1983. Method for biological quality assessment of watercourses in Belgium. Hydrobiologia 100: 153–183.

De Pauw N. and Vannevel R. 1991. Macroinvertebrates and Water Quality. Stichting Leefmilieu, Antwerp, Belgium, 316 p. (in Dutch).

De Pauw N. and Heylen S. 2001. Biotic index for sediment quality assessment of watercourses in Flanders, Belgium. Aquat. Ecol. 35(2): 121–133.

EU 2000. Directive of the European Parliament and of the Council 2000/60/EC Establishing a Framework for Community Action in the Field of Water Policy. European Union, The European Parliament, The Council, PE-CONS 3639/1/00 REV 1 EN, 62p. + annexes.

Foody G.M. 1999. Applications of the self-organising feature map neural network in community data analysis. Ecol. Model. 120: 97–107.

Furse M.T., Moss D., Wright J.F. and Armitage P.D. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain on the prediction of macroinvertebrate communities. Freshwater Biol. 14: 257–280.

Giraudel J.L. and Lek S. 2003. Ecological applications of unsupervised neural networks. In: Recknagel F. (ed.), Understanding Ecology by Biologically Inspired Computation. Springer-Verlag, Berlin Heidelberg, pp. 15–33.

Goethals P.L.M. and De Pauw N. 2001. Development of a concept for integrated ecological river assessment in Flanders (Belgium). J. Limnol. 60: 7–16.

Halkidi M., Batistakis Y. and Vazirgiannis M. 2001. On clustering validation techniques. J. Intell. Inf. Syst. 17(2–3): 107–145.

Hawkes H.A. 1979. Invertebrates as indicators of river water quality. In: James A. and Evison L. (eds), Biological Indicators of Water Quality. John Wiley, Chichester, UK.

Hawkins C.P., Norris R.H., Gerritsen J., Hughes R.M., Jackson S.K., Johnson R.K. and Stevenson R.J. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. J. N. Am. Benthol. Soc. 19: 541–556.

Hildrew A.G. 1992. Food webs and species interactions. In: Calow P. and Petts G.E. (eds), The Rivers Handbook: Hydrological and Ecological Principles, Vol. I. pp. 309–329.

Hill M.O. 1973. Reciprocal averaging: an eigenvector method of ordination. J. Ecol. 61: 237–249.

Hill M.O. 1979. DECORANA – a FORTRAN Program for Detrended Correspondence Analysis and Reciprocal Averaging. – Ecology and Systematics. Cornell University, Ithaca, New York, USA, 52 p.

IBN 1984. Norme Belge T 92-402. Biological Water Quality. Determination of a Biotic Index Based on Aquatic Macroinvertebrates. Institut Belge de Normalisation, Brussels, Belgium, 11p.

Jackson D.A. 1993. Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association and ordination methods. Hydrobiologia 268: 9–26.

Kohonen T. 1982. Self-organization and associative memory. Springer-Verlag, Berlin, Germany, 312p.

McIntosh R.P. 1967. The continuum concept of vegetation. Bot. Rev. 33: 130–187.

Palmer A. and White P.S. 1994. On the existence of ecological communities. J. Veg. Sci. 5: 279–282.

Palmer A., Ambrose R.F. and LeRoy Poff N. 1997. Ecological theory and community restoration ecology. Restor. Ecol. 5(4): 291–300.

Pardo I. and Armitage P.D. 1997. Species assemblages as descriptors of mesohabitats. Hydrobiologia 344: 111–128.

Parsons M., Thoms M.C. and Norris R.H. 2003. Scales of macroinvertebrate distribution in relation to the hierarchical organization of river systems. J. N. Am. Benth. Soc. 22(1): 105–122.

Prati L., Pavanello R. and Pesarin F. 1971. Assessment of surface water quality by a single index of pollution. Water Res. 5: 741–751.

Rosenberg D.M. and Resh V.H. 1993. Introduction to freshwater biomonitoring and benthic macroinvertebrates. In: Rosenberg D.M. and Resh V.H. (eds), Freshwater Biomonitoring and Benthic Macroinvertbrates. Chapman and Hall, New York, USA.

Ruse L.P. 1996. Multivariate techniques relating macroinvertebrate and environmental data from a river catchment. Water Res. 30(12): 3017–3024.

Schneiders A. and Verheyen R. 1998. A concept of integrated water management illustrated for Flanders (Belgium). Ecosyst. Health 4(4): 256–263.

Schneiders A., Wils C. and Verheyen R. 1999. The use of ecological information in the selection of quality objectives for river conservation and restoration in Flanders (Belgium). Aquat. Ecosyst. Health Manage. 2: 137–154.

Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. Det Kong Danske Vidensk Selsk Biol Slr (Copenhagen) 5(4): 1–34.

Ter Braak C.J.F. and Smilauer P. 1998. Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4). Microcomputer Power, Ithaca, NY, USA, 352p.

Ter Braak C.J.F. and Verdonschot P.F.M. 1995. Canonical correspondence analysis and related multivariate analysis in aquatic ecology. Aquat. Sci. 57(3): 255–289.

Townsend C.R. 1989. The patch dynamic concept of stream community ecology. J. N. Am. Benthol. Soc. 8: 36–50.

Ultsch A. and Siemon H.P. 1990. Kohonen's self organizing feature maps for exploratory data analysis. In: Proc. INNC'90, Int. Neural Network Conf., Kluwer, Dordrecht, The Netherlands, pp. 305–308.

Vandenberghe V., Goethals P.L.M., Van Griensven A., Meirlan J., De Pauw N., Vanrolleghem P. and Bauwens W. 2004. Application of automated measurement stations for continuous water quality monitoring of the Dender river in Flanders, Belgium. Environ. Monit. Assess., in press.

Vannote R.M., Minshall G.W., Cummins K.W., Sedell J.R. and Cushing E. 1980. The river continuum concept. Can. J. Fish Aquat. Sci. 37: 130–137.

Van Tongeren O. 1986. Flexclus, an interactive program for classification and tabulation of ecological data. Acta Bot. Neerl. 35(3): 137–142.

Verdonschot P.F.M. 1990. Ecological Characterization of Surface Waters in the Province of Overijssel (the Netherlands). PhD thesis, Wageningen, 255p.

Verdonschot P.F.M. 1995. Typology of macrofaunal assemblages: a tool for the management of running waters in the Netherlands. Hydrobiologia 297: 99–122.

Verdonschot P.F.M. 2000. Integrated ecological assessment methods as a basis for sustainable catchment management. Hydrobiologia 422/423: 389–412.

Verdonschot P.F.M. and Nijboer R.C. 2000. Typology of macrofaunal assemblages applied to water and nature management: a Dutch approach. In: Wright J.F., Sutcliffe D.W. and Furse M.T. (eds), Assessing the Biological Quality of Fresh Water: RIVPACS and Other Techniques. Ambleside, UK, FBA, pp. 241–262.

Verdonschot P.F.M., Nijboer R.C., Janssen S.N. and van den Hoorn M.W. 2000. Ecological typology limburg. Alterra, Wageningen, The Netherlands, 78p. (in Dutch).

Verdonschot P.F.M. and Nijboer R.C. 2002. Towards a decision support system for stream restoration in the Netherlands: an overview of restoration projects and future needs. Hydrobiologia 478(1–3): 131–148.

Vesanto J., Himber J., Alhoniemi E. and Parhankangas J. 2000. SOM toolbox for MATLAB 5. Helsinki University of Technology, Publications in Computer and Information Science, Report A57, Helsinki, Finland, 59p.

VMM 1997. Water quality 1996. Report Surface Water Monitoring. Aalst, Belgium (in Dutch).

Walley W.J., Martin R.W. and O'Connor M.A. 2000. Self-organising maps for classification of river quality from biological and environmental data. In: Denzer R., Swayne D.A., Purvis M. and Schimak G. (eds), Environmental Software Systems: Environmental Information and Decision Support. IFIP Conference Series, Kluwer Academic Publishers, pp. 27–41.

Walley W.J. and O'Connor M.A. 2001. Unsupervised pattern recognition for the interpretation of ecological data. Ecol. Model. 146: 219–230.

Whittaker R.H. 1967. Gradient analysis of vegetation. Biol. Rev. 49: 207–264.