Spatial and temporal patterns of benthic invertebrate communities in an intermittent river (North Africa)

A. Arab¹, S. Lek², A. Lounaci³, Y. S. Park²*

¹ Université des Sciences et de la Technologie Houari Boumediène, Faculté des Sciences Biologiques, BP 39, El Alia, Alger, Algeria
² Laboratoire Dynamique de la Biodiversité (LADYBIO), UMR 5172, CNRS - Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France
³ Laboratoire d’Hydrobiologie, Faculté des Sciences Biologiques & des Sciences Agronomiques, Département de Biologie, Université Mouloud Mameri de Tizi-Ouzou, Algeria.

The spatial and temporal distribution patterns of benthic macroinvertebrates were studied in an intermittent river in Algeria (Chelif wadi, North Africa), by using the self-organizing map (SOM), an unsupervised artificial neural network. The samples were collected monthly at 8 sampling sites (630 to 20 m above sea level) and community variation was analysed in space and time. Overall, the study sites showed a very poor macroinvertebrate fauna: more than 60% of samples contained less than 11 species, and 99% had less than 30 species. Furthermore, most species displayed very low abundance: 66% of the species were represented by less than 20 individuals (abundance). Among the identified taxa, Chironomidae was the dominant taxon at all sampling sites except at the most upstream site (630 m a.s.l.) where it was replaced by Coleoptera. Concerning monthly changes, the species richness was very low in August and October. Through the learning process of the SOM, samples were classified into four clusters by the SOM, and the classification was mainly related to the location of the sampling sites. Benthic macroinvertebrates were divided into four classes, which revealed the influence of pollution on their longitudinal distribution in this stream. According to the distribution gradients of the environmental variables on the SOM map, their influence on the classification of the sampling sites could be assessed effectively.

Keywords: macroinvertebrates, intermittent stream, species richness, self-organizing map, artificial neural network, North Africa.

Introduction

The distribution of benthic macroinvertebrates is largely related to environmental factors (Richards et al. 1993, Tate & Heiny 1995, Benbow et al. 2003), and their communities change both spatially and temporally (Hynes 1970, Townsend & Hildrew 1994). Changes in the benthic community structure were often assessed by examining taxonomic compositions along the stream continuum (Grubaugh et al. 1996, González et al. 2003). Spatial variations have been attributed to habitat characteristics (e.g., Brown 1984, Brown et al. 1995, Boyero 2003), but also to stochastic temporal fluctuations of the environment (Ives & Kopfer 1997, McIntosh et al. 2003). Based on the dynamic equilibrium concept (Curry 1972), streams form a balance between physical parameters (such as width, depth, velocity, and sediment load), and biological factors. Temporal variations in community assemblages can result from habitat seasonality, organism phenology, and/or disturbance. Flooding represents the dominant type of natural disturbance along most river corridors (Welcomme 1979, Puckridge et al. 1998, Arscott et al. 2003) and it is likely that all stream ecosystems are disturbed by fluvial forces to some degree (Reice 1985, Statzner et al. 1987, Ward 1989). Changes in hydraulic stress due to tributary confluences or changes in channel geometry have also been suggested to influence
longitudinal patterns of invertebrate diversity (Statzner & Higler 1986, Arscott et al. 2003). Finally, thermal patterns in rivers are also thought to be a major factor structuring benthic macroinvertebrate communities (Vannote & Sweeney 1980, Ward 1985, Arscott et al. 2003).

The Mediterranean basin is characterized by an extensive range of physical conditions, including severe flooding and droughts (Pires et al. 1999). The thermal variations are sometimes brutal, and the irregular and violent precipitations are concentrated over short periods. Mediterranean rivers are characterized by the irregularity of the flows and brutal hydrological fluctuations. The annual flow mode is marked by a maximum in spring and autumn, and a severe low water level in summer (Giudicelli et al. 1985).

In arid and semi-arid regions, intermittent streams are the dominant type of streams (Comin & Williams 1994), and are ubiquitous in areas with Mediterranean type climates (Gasith & Resh 1999). The drying process poses several physico-chemical and biological challenges to the biota of intermittent streams (Williams 1996). The patterns of water disappearance are of fundamental importance, whether they are predictable (i.e., part of a stable cycle) or unpredictable (i.e., linked to the vagaries of local climate) (Williams & Hynes 1976, Wiggins et al. 1980, Delucchi 1988). The length of the dry period thus influences the decrease in habitat size. Temperature changes result in subtle changes in water density which may affect the smaller instars of insects, and ionic concentration tends to rise significantly as temporary water dries out, which may affect pH. Oxygen will be further affected by changes in water temperature which will increase as the water disappears in summer-dry water bodies (e.g., autumnal ponds, Wiggins 1973), while it will decrease in winter-dry bodies (e.g., aestival ponds, Daborn & Clifford 1974).

The running water ecosystems of Northern Algeria (Mediterranean basin) are physically very diverse, not only between geographical areas but also between upstream and downstream sections. In this area, the dominating feature of the rivers is the irregularity of the flows. Floods and low water levels are the two major events to which these lotic ecosystems are subjected, so temporary biotopes are largely dominant. In Algeria, aquatic organisms are severely affected by the aestival drought. Most early works on the hydrosystems of North Algeria were mainly devoted to the description of species, but few concerned their ecology or biogeography (Reiche 1869, Seurat 1922, Navas 1929). More recently, some studies have been conducted on the fauna of inland waters of this country (Gagneur et al. 1986, Arab 1989, Lounaci et al. 2000). The objective of this work was thus to determine the spatial and temporal patterns of benthic macroinvertebrate assemblages in an intermittent stream in North West Algeria.

Materials and methods

Data collection

Benthic macroinvertebrates were collected monthly at 8 sampling sites in the Chelif river, North West Algeria (Fig. 1), from December 1989 to December 1992. The Chelif is the wadi which drains part of the high plateau, and it is one of the longest rivers in North Africa (length : 700 km). The Chelif plain receives only moderate, undependable rainfall (average : 400 mm annually), and evaporation is intense. The lower reaches of the river basin are, however, cultivated with the aid of irrigation. Three main dams have been constructed on the Chelif system at Ksar el-Boukhari (1932), Ech-Chelif (El-Asnam; 1932), and El-Khemis (1939). The Chelif drainage basin covers a total area of 44 630 km². The altitude of the sampling sites ranged from 630 m (St1) to 20 m a.s.l. (St8). Generally, an aestival absence of flow occurs from May to November, if there is not enough rainfall. Especially, sites St4 and St8 were highly influenced by the lack of flow. The general characteristics of the physical factors are summarized in Table 1.

Macroinvertebrates were collected using a Surber sampler (30 cm x 50 cm, mesh size 250 µm). At each sampling time, 6 samples were collected as replicates. The samples were preserved in 5 % formalin and in-

![Fig. 1. Map of the study area, and location of the sampling sites.](image-url)
vertebrates were identified to species or morphospecies level. Due to the low abundance of the fauna in the study areas, the samples collected over three years were summarised monthly at each sampling site and used in the data analysis. Substrate composition (%) at each sampling sites was measured by direct observation: pebble (25-50 mm), gravel (5-25 mm), sand (0.06-5 mm), silt (<0.06 mm), and organic matter.

Community assemblages

To evaluate the differences in community structures between sampling sites from upstream to downstream areas, community indices such as species richness and Shannon diversity index were evaluated at each sampling site. A Kruskal-Wallis test was conducted to test significant differences of species richness between clusters defined by a self-organising map (see below). Mann-Whitney tests were conducted to evaluate significant differences of community indices between sites. Statistical analyses were carried out using the statistical software STATISTICA (StatSoft 2002).

Modelling procedure

The dataset of macroinvertebrate assemblages was applied to an adaptive learning algorithm, the self-organizing map (SOM, unsupervised neural network), to bring out distribution patterns of communities in space and time. The species abundances were scaled between 0 and 1 in the range of the minimum and maximum values within a species, after a log-transformation process in order to reduce variations in abundance. The SOM approximates the probability density function of the input data (Kohonen 2001). The SOM has found wide applications in the fields of data exploration, data mining, data classification, data compression, and biological modelling, due to its properties of neighbourhood preservation and local resolution of the input space proportional to the data distribution. The SOM consists of input and output layers connected with computational weights (connection intensities). The array of input neurons (i.e., computational units) operates as a flow-through layer for the input vectors, whereas the output layer consists of a two-dimensional network of neurons arranged in a hexagonal lattice. In the learning process of the SOM, the biological data were initially subjected to the learning network. Then, the weights were trained for a given dataset of the network, each neuron \( k \) of the output layer computes the summed distance between weight vector \( w \) and input vector \( x \). The output layer consists of \( N \) output neurons (i.e., computational units, \( 24 = 6 \times 4 \) in this study) which usually constitute a 2D grid for better visualization. The output neurons are considered as virtual units to represent typical patterns of the input dataset assigned to their units after the learning process. Among all \( N \) virtual units, the best matching unit (BMU) which has the minimum distance between weight and input vectors becomes the winner. For the BMU and its neighbourhood units, the new weight vectors are updated by the SOM learning rule. This results in training the network to classify the input vectors by the weight vectors they are closest to. The virtual assemblages can be obtained by denormalizing weight vectors trained in the learning process. To define clusters between virtual units of the SOM map, a hierarchical cluster analysis was used with Ward’s linkage method with Euclidean distance measure. Based on linkage distances, the map was classified into several groups indicating differences of assemblages at different scales. During the learning process of the SOM, the map units that are topographically close in the array will activate each other to learn something from the same input vector. The weight vectors tend to approximate the probability density function of the input vector. Therefore, the
visualization of these vectors is convenient to understand the contribution of each input variable to the clusters on the trained SOM (Kohonen 2001, Park et al. 2003a). To analyse the contribution of variables to cluster structures of the trained SOM, the value of each input variable (component) calculated during the training process was visualized in each unit on the trained SOM map on a grey scale.

Results

Community assemblage

Overall, 175 species were identified from the various samples. Nevertheless, more than 60% of the samples contained less than 11 species, and 99% had less than 30 species (Fig. 2a). In addition, most species displayed very low abundance: 66% of the species were represented by less than 20 individuals (Fig. 2b).

Among the identified taxa, Chironomidae was the dominant taxon at all sampling sites, except site St1 where Coleoptera was dominant (Table 2). The taxonomic richness of macroinvertebrates showed fluctuations along the Chelif course. It was the highest at site St2 (71 species), decreased at the downstream site St3 (54 species), and showed the lowest value (14 species) at site St4 where water was heavily polluted. Downstream of site St4, the taxonomic richness increased gradually. Sampling sites St1-3 showed relatively high species richness, whereas sampling site St4 showed

Table 2. Species richness in different taxa at the various sampling sites.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>St1</th>
<th>St2</th>
<th>St3</th>
<th>St4</th>
<th>St5</th>
<th>St6</th>
<th>St7</th>
<th>St8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ephemeroptera</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>17</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Oligochaeta</td>
<td>2</td>
<td>9</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Odonata</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Diptera*</td>
<td>10</td>
<td>17</td>
<td>11</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Chironomidae</td>
<td>5</td>
<td>21</td>
<td>13</td>
<td>3</td>
<td>11</td>
<td>10</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>71</td>
<td>54</td>
<td>14</td>
<td>37</td>
<td>35</td>
<td>51</td>
<td>57</td>
</tr>
</tbody>
</table>

* Diptera except Chironomidae
Considering monthly changes, the species richness was very low around September (August and October). During this period, no species were collected at sampling sites St1, 2, 4-6, 10. The species richness was the highest in June at most sampling sites. The Shannon diversity index which takes account of the specific richness as well as the abundances showed the same patterns (Fig. 4a). The index showed the highest value at the upstream site St1, and it decreased progressively to site St4 (lowest values). The Shannon index then increased gradually from site St5 to site St8, although there were some variations. At downstream areas, the differences were not significant (Mann-Whitney test, p>0.05) (Table 3).

### Community distribution patterns

Macroinvertebrate communities were patterned on the 2D lattice map through the leaning process of the SOM with abundance of species (Fig. 5a). Overall, communities were classified according to the location of the sampling sites and monthly densities, although large variations were observed. For example, most samples from site St1 were located in the middle-left areas of the SOM, and samples from mid-downstream areas (i.e., St4-8) were grouped in the upper right areas of the SOM map, whereas samples from summer and autumn were grouped together in the upper-left areas.

<table>
<thead>
<tr>
<th>Site</th>
<th>St2</th>
<th>St3</th>
<th>St4</th>
<th>St5</th>
<th>St6</th>
<th>St7</th>
<th>St8</th>
</tr>
</thead>
<tbody>
<tr>
<td>St1</td>
<td>0.526</td>
<td>0.074</td>
<td>0.012*</td>
<td>0.042*</td>
<td>0.0178*</td>
<td>0.325</td>
<td>0.213</td>
</tr>
<tr>
<td>St2</td>
<td>0.066</td>
<td>0.004**</td>
<td>0.057</td>
<td>0.398</td>
<td>0.895</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>St3</td>
<td>0.031*</td>
<td>0.46</td>
<td>0.712</td>
<td>0.166</td>
<td>0.644</td>
<td></td>
<td></td>
</tr>
<tr>
<td>St4</td>
<td>0.09*</td>
<td>0.017*</td>
<td>0.006**</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St5</td>
<td>0.533</td>
<td>0.074</td>
<td>0.295</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St6</td>
<td>0.268</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St7</td>
<td>0.386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.01
of the SOM. Samples collected at sites St1, 2, 4-6, 8 in August-October were grouped together and distinguished from others in the upper-left areas of the SOM map. These samples were characterised by low species richness.

A hierarchical cluster analysis with a Ward linkage method was used to define cluster boundaries among the SOM units (Fig. 5b). Samples in the SOM map were further classified into four groups (I-IV) according to their community similarities. Cluster I was characterised by samples from downstream areas, cluster II was identified by samples from site St1, cluster III was characterized by samples from sites St3 and St8, and cluster IV was characterized by site St3. Cluster I showed the lowest species richness and significantly differed from other clusters (Kruskal-Wallis test, p<0.001), whereas there were no significant differences between clusters II, III, and IV (Fig. 5c).

After the SOM learning process, estimated species profiles were obtained from each SOM unit, and summarized in each cluster. Through these profiles, we were able to evaluate the relative importance of each species in determining the community patterns in the SOM map. Figure 6 shows typical distribution patterns of species in each cluster. The distribution pattern of each species on the SOM units was assigned to one of four patterns based on the occurrence cluster of the highest abundance areas. Cluster I was mainly characterised by *Limnodrilus hoffmeisteri*, cluster II was dominated by *Micronecta scholtzi* and *Panarais* sp., cluster III by *Baetis neglectus*, *Cloeon gr. dipterum*, and *Cri cotopus bicinctus*, and cluster IV was characterised by *Baetis pavidus*, *Caenis luctuosa*, *Rheocricotopus chalybeatus*, and *Tubifex tubifex*.

**Differences of substrate composition**

Substrate composition of sampling sites were visualised in different clusters defined by the SOM trained with invertebrate data (Fig. 7). Silt showed high values in clusters I and II, and low values in clusters III and IV. Pebbles and gravels had high values in cluster III. In cluster IV, there were no significant differences between substrate types, although sand showed the highest value. All substrate types excluding organic matter were significantly different between clusters (Kruskal-Wallis test, p<0.001).
Discussion

Highly heterogeneous ecosystems affect the relationship between spatial and temporal variation in environmental constraints and the responses by organisms to them (Milne 1991). Numerous studies have focused on the biological implications of these constraints in terms of processes and mechanisms (Palmer et al. 1995, Palmer & Poff 1997, Townsend et al. 1997). The community structure is changed by perturbations in the environment and the degree of structural change is used to assess the intensity of the environmental stress (Hellawell 1986). Species richness is influenced by environmental conditions (Legendre & Legendre 1998). A stable environment contains more species and more niches, because a more stable environment involves a higher degree of organization and complexity of the food web (Margalef 1958). The
niches, the species richness of Diptera (with high Chironomidae richness) was relatively high at most of our sampling sites, except at the upstream site St1, probably because of a pollution gradient. During the field survey, we evaluated the water quality of sampling sites according to the suspended materials and the colour of water, ranging from 1 (very polluted, dark water) to 5 (relatively clean, transparent). Site St1 was the cleanest areas (5), whereas site St4 had the worst water quality (1). Water quality at other sites ranged between the estimates for these two sites. Species colonization of intermittent streams is dependent on environmental factors such as drying conditions, and on behavioural and/or physiological adaptations to habitat drying. Tolerance of species to drying by behavioural and/or physiological adaptations is the basis for studies that have concluded that the assemblages found are comparable to those of nearly perennial streams (Williams 1987). With the loss of water, aquatic insects counter the ‘terrestrialisation’ of their habitat by means of physiological tolerance and migration. Although the adaptations shown by individual species are varied, members of a same major taxonomic groups tend to have similar stages in their life cycles that resist drying out. Thus, mayflies, lepidos, and mosquitoes largely survive drought as eggs (Lehmkuhl 1973, Wood et al. 1979), many beetles and hemipterans survive as adults (Macan 1939), and stoneflies survive by diapausing as early instar nymphs (Harper & Hynes 1970). Chironomids may diapause as larvae or eggs (Thienemann 1954). Pires et al. (2000) reported that Diptera was the dominant taxa in intermittent streams in Portugal and concluded that Diptera dominated the fauna because they are more tolerant to drought conditions and have more efficient recolonization mechanisms. Among the study sites, sites St1-3 showed the highest Shannon diversity (Jørgensen & Padisak 1996).
ter supply was insufficient at this site to provide habitats to aquatic organisms, and for self-purification. Site St5 was also influenced by industrial waste and showed low values of Shannon diversity and species richness. However, the indices increased in downstream areas (St6-8), and became similar to those calculated at sites St1-3. This pattern suggests that water quality was improved towards downstream reaches, due to self-purification.

The SOM showed its ability to classify input vectors as well as to visualize relationships among input variables in their contribution to the classification (i.e., Chon et al. 1996, Park et al. 2001, 2003a, 2003b, 2003c, 2004, Giraudel and Lek 2001, Céréghino et al. 2003). A clear gradient of a variable in the SOM map represents a high contribution gradient to the classification of input vectors. In this study, the samples in space and in time were classified into four clusters (Fig. 5). According to the distribution gradients of species on the SOM map, samples as well as species could be assessed effectively. The main property of the SOM is the dimension reduction achieved by the learning process. High-dimensional input data is mapped into low dimensions (commonly a 2-dimensional map), with information on the relationships between input signals preserved as much as possible. Through this process, neighbouring input values are mapped onto neighbouring (or the same) nodes according to some metric defined in the output space and neighbourhood topology is preserved. Additionally, the SOM averages the input dataset in weight vectors through the learning process and thus removes noise (Vesanto et al. 1998).

The averaging effect was implemented in the first step of the combined model in this study. It is well known that classical clustering techniques are sensitive to the problem of outliers. Each outlier takes its place in one unit of the map, and only the weights of that unit and its nearest neighbours are affected. There is no effect on the other units. Likewise, scattered nodes on the map suggest the presence of an outlier (Lek & Guégan 2000).

The effects of drought were clearly reflected by monthly changes of communities. The species richness was very low around September (August and October, i.e. the dry season). These characteristics were also well implemented in the SOM map. Samples of drought periods were located in the upper left areas of the SOM map, showing low species richness. Meanwhile, samples showing relatively high species richness were located in the lower areas of the SOM map. Clusters defined by the SOM showed different habitat conditions at the sampling sites. Cluster I (i.e., sites St4-7) and cluster II (site St1) were characterised by high proportion of silt, whereas cluster III (sites St2, 5, 7, 8) was characterised by relatively high coverage of pebbles and gravels. According to these differences of substrate composition, habitat suitability for species can be suggested. According to the species distribution in the SOM map in this study, cluster I was characterised mainly by Limnodrilus hoffmeisteri, cluster II was dominated by Micronecta scholtzi and Panarais sp., cluster III by Baetis neglectus, Cloeon gr. dipterum, and Cricotopus bicinctus, and cluster IV was characterised by Baetis pavidus, Caenis luctuosa, Rheocricotopus chalybeatus, and Tubifex tubifex. However, the distributions of species are also governed by many different environmental parameters. In particular, intermittent streams are severely influenced by flooding and drought. To determine macroinvertebrate abundances in intermittent streams of Portugal, substrate was not important factor, but temperature, conductivity, and drought were the most important factors (Pires et al. 2000).

In summary, the benthic macroinvertebrate fauna collected in the study intermittent river was very poor. Chironomidae was the dominant taxa, and species richness was very low around September, during the dry season. Through an adaptive learning algorithm, benthic macroinvertebrates were patterned and sampling sites were classified into four clusters according to similarities in their community composition. The clusters were mainly related to the downstream location of the sampling sites, and to the presence of the above-mentioned dominant species. Finally, our results showed that the SOM approach provides an efficient characterization of the distribution patterns of aquatic organisms in the domains of space and time.

Acknowledgements

We are grateful to two anonymous reviewers for their insightful comments on an earlier draft of this manuscript.

References


