Comparative community analysis of benthic macroinvertebrates and microorganisms across different levels of organic pollution in a stream by using artificial neural networks

Mi-Young Song, Se-Eun Lee[1], Jungim Park, Jieun Park, Byunghyuk Kim[1], Seungchul Koh[1], Keunsup Lee, Young-Seuk Park[2] and Tae-Soo Chon*
Division of Biological Sciences, Pusan National University, Busan 609-735 Korea
[1]Division of Civil and Environ Engineering, Korea Maritime University, Busan 606-791 Korea
[2]Department of Biology, Kyung Hee University, Seoul 130-701 Korea
miysong@pusan.ac.kr, amsaza777@hotmail.com, jipark20@hanmail.net, jumpje@hanmail.net, nomdj@hotmail.com, skoh@mail.hhu.ac.kr, klee@pusan.ac.kr, parkys@khu.ac.kr, *tschon@pusan.ac.kr (Correspondence author)

Abstract: - Macroinvertebrates and microorganisms were sampled from the same habitats in a stream across different levels of pollution. Microorgansims collected from the field were identified with Denature Gradient Gel Electrophoresis (DGGE). Through learning process of Self-Organizing Map (SOM), community patterns were comparatively revealed in a separate and integrative manner for two taxa. Community of macroinvertebrates patterned by SOM showed the gradient of organic pollution from clean to polluted states, while the impact of natural disturbance (e.g. flooding in summer) was also disclosed on the patterned map. In contrast, patterns on SOM based on microorganisms responded to local differences at the sample sites more, although overall differences in polluted and clean sites were still produced. The inter-taxa community patterning was efficient in integrating community patterns of two taxa. While the gradient was similarly revealed from the clean to polluted sample sites, clustering and subclustering were more specifically associated with local differences in the sample sites. On the integrated map, spectrum of abundance was also revealed for two taxa. Species in macroinvertebrates appeared to be more associated with clean sites, while species in microorganisms were more related to polluted sites. The species appearing in different clusters were further associated with differences in various environmental factors. The inter-taxa presentation of community data assisted by techniques in ecological informatics such as SOM could be an efficient means of the integrative assessment in aquatic ecosystems.

Key-Words: - Macroinvertebrate, Microorganism, DGGE, Polluted stream, SOM, Integrative analysis

1 Introduction

In polluted conditions, macroinvertebrates have been widely used for assessing water quality and ecological status of aquatic systems with various advantages of taxonomic diversity, sedentary in behaviors and long life cycles [1]. As consumers in aquatic ecosystems, benthic macroinvertebrates serve as a link between producers (e.g. algae) and decomposers (e.g. microorganisms) in a food chain, and maintain a diverse spectrum in species composition. While species richness and diversity are high in clean conditions, only a few tolerant species are selectively dominant in polluted conditions (e.g. [1], [2]). In this regard, numerous indicators have been developed from benthic macroinvertebrate communities to assess water quality such as BMWP, EPT, RIVPACS, etc. (e.g. [2], [3], [4]). Along with benthic macroinvertebrates, microorganisms also play an important role in aquatic systems regarding decomposition of organic matters [5], and energy recovery in the high-level food chain through microbial loop [6]. Consequently microorganisms sensitively respond to *in situ* availability of organic matters in polluted conditions and are considered as an efficient group to indicate water quality [7].

If microorganisms and macroinvertebrates are analyzed in an integrative manner, the ecological status of aquatic ecosystems would be comprehensively revealed from both aspects of decomposition and consumption. There have been numerous accounts of research on community-habitat relationships and water quality assessment separately for microorganisms (e.g. [8], [9], [10]) and for

macroinvertebrates (e.g. [11], [3], [4]). However, microorganisms and macroinvertebrates have been seldom compared directly at the same sample sites except for a few cases. Recently Yeager et al [12] checked microbial community structure and function in response to larval chironomid feeding pressure in a microcosm and reported a marked difference in microbial community corresponding to increase in midge population density. Due to difficulty in characterizing microorganisms such as identification and measurement of *in situ* population size [13], interactions between microorganisms and macroinvertebrates have been ascertained from more concentrated, but limited, scopes mainly regarding the effects of macroinvertebrates on bacterial colonization of biofilms ([14], [15]) or on litter processing ([16], [17], [18], [19]). There have been no integrative inter-taxa community analyses specifically related to changes in water quality. In this study we intend to directly compare two taxa through "community by community" basis in order to reveal distribution ranges of two taxa across different levels of pollution.

In order to analyze the complex datasets for the inter-taxa communities, we utilized Self-Organizing Map (SOM) [20]. In ecology, techniques in ecological informatics such as artificial neural networks have been demonstrated in patterning community data (e.g. [21], [22]). Recently artificial neural networks have been additionally used for molecular biological study including patterning the DNA microarray data [23], and bacterial genome [24].

2 Materials and Methods

2.1 Study sites

We selected 5 sample sites across different levels of organic pollution (clean (DUK), polluted (DKH, DDK, and DKS) and recovering (DAG)) in the Daechon Stream, Busan, Korea (Fig. 1), and surveyed microorganisms and macroinvertebrates in different seasons (January, April, July and October) from January 2002 to April 2003. Along with sampling, chemical factors such as BOD_5, conductivity, inorganic nutrients (e.g. ammonium, nitrate, and phosphate in sediment and water column) and DOC (Dissolved Organic Carbon) were measured according to the Standard methods [25].

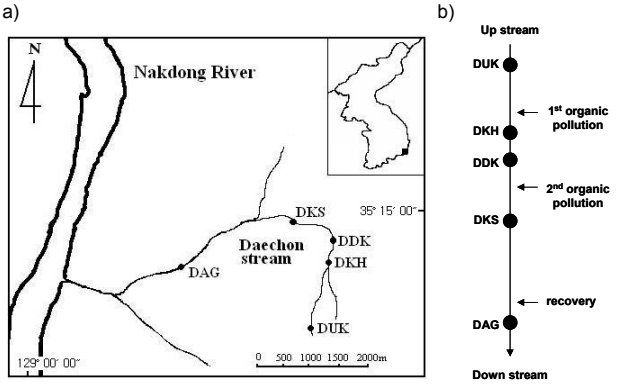

Fig. 1. Location of the sampling sites in the Daechon Stream, Busan, Korea. Site map (a), and pollution input in the stream (b).

2.2 Sampling and molecular biological analysis of microorganisms

Sampling of microorganisms was carried out on the same sample sites before benthic macroinvertebrates were collected. The samples were collected from the surface of sediments using a disposable syringe (100 *ml*) with its injection portion cut off. Denature Gradient Gel Electrophoresis (DGGE) was utilized to identify DNA sequences [9]. The data for the sequences of the output bands were patterned by using SOM. The flow chart of microbial community analysis was shown in Fig. 2.

2.2.1 Nucleic acid extraction

Total DNA extractions were made for each sediment sample (0.5g) using Fast DNA SPIN Kit (Q-BIO Gene). The DNA preparations were used as template DNA's in the subsequent PCR-DGGE. Universal primers 27F and 1492R [26] targeting eubacterial 16S rDNA were utilized for a first round PCR to circumvent multiple amplification. The PCR condition was: denaturation (95°C, 5min), 30 cycles of the standard PCR (95°C, 30s; 55°C, 30s; 72°C, 1.5min), and a final chase reaction of (72°C, 7min). A second PCR was subsequently performed using another primer set 341F-GC and 518R targeting the V3 region of the 16S rDNA [26]. The subsequent PCR was a touchdown PCR using annealing temperatures spanning 65-55°C.

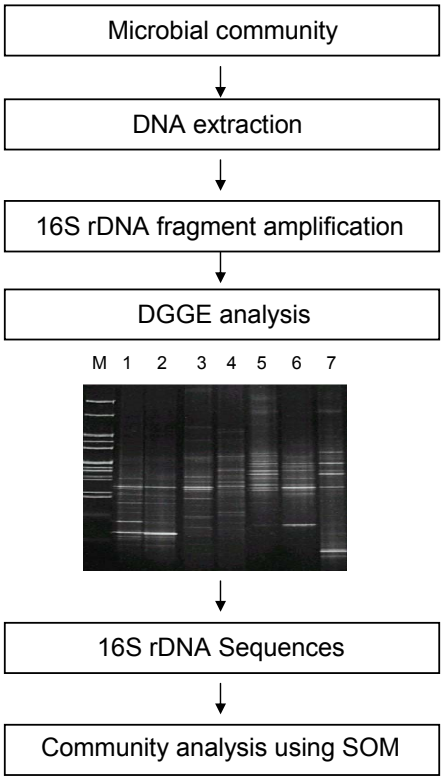

Fig. 2. The procedure for analysis, identification and community grouping of microorganisms collected across different levels of pollution in aquatic conditions.

2.2.2 DGGE and image analysis

The PCR products of the V3 region of the 16S rRNA were subjected to DGGE to separate each amplicon according to its DNA sequence difference (Fig. 2). The DGGE was performed at 60°C using D-Code System (Bio-Rad) according to the manufacturer's instruction. Gradient of denaturation within gel ranged from 30 to 60% [26]. The gels were run at 60V for 16h. The bands separated in the DGGE gels were digitized using an image analysis system (Model Kodak, USA). The values from the digitized images were stored as Excel files, and were regarded as population densities.

2.2.3 Identification of microbial communities based on 16S rRNA sequence analysis

Representative dominant DNA fragments separated within the DGGE gels were cut out from the gels and eluted into water and amplified again using the primer set 341F and 518R. The products (100 μl) were purified using PCR purification kit. Each DNA fragment purified from the DGGE profiles was reamplified and purified and then sequenced. The DNA fragment was considered as a eubacterial population and identified up to genus or species level compared with the known DNA sequences through BLAST search and the sequence alignment at NCBI. The bacterial population densities which were identified as the same genus or species were pooled and considered as the same group for pattern analysis of microbial communities (Fig. 2).

2.3 Sampling of macroinvertebrates

At each sample site, macroinvertebrates were collected with the Surber sampler (30 cm × 30 cm, 500 μm mesh; [25]) in approximately 10 cm depth. In the laboratory the invertebrate specimens were sorted, identified and counted under microscopes. Identification was based on Merritt and Cummins [27], Pennak [28] and Brinkhurst [29].

2.4 Community indices

In order to reveal community structure of the sample sites, Species Richness (SR), Diversity (H') and Evenness (J'), and Density were estimated for both microorganism and macroinvertebrate communities [1]. SR indicates the total number of species collected in a sample, while Shannon Diversity was measured as:

$$H' = -\sum_{i=1}^{k} p_i \log p_i \qquad (1)$$

Here, k is the number of species and p_i is the proportion of the observations found in species i. Evenness is calculated as follows:

$$J' = \frac{H'}{H'_{\max}} \qquad (2)$$

Here, H'_{max} is the maximum possible diversity for a set of data consisting of k species.

2.5 Data analysis

The similar number of species was assigned as input variables for the data of macroinvertebrates and microorganisms. Communities for two taxa were scaled between 0 and 1 in the range of the minimum and maximum values and were used for input data for training with SOM. The community data were log-transformed before rescaling process for each taxon. After patterning was carried out for microorganisms and macroinvertebrates separately by SOM, the integrative analyses were additionally conducted for the combined data of microorganisms and macroinvertebrates. The datasets of community abundance for microorganisms and

macroinvertebrates were merged to be used as input for training with SOM.

In SOM a linear array of $M \times N$ artificial neurons (i.e., computation nodes), with each neuron being represented as j and k, is arranged in two dimensions (Fig. 3). Suppose a community data containing P species, and the density of species, i, is expressed as a vector x_i. The vector x_i is considered to be an input layer to the SOM. In the network each neuron, (j, k), is supposed to be connected to each node, i, of the input layer. The connectivities are represented as weights, $w_{ijk}(t)$, adaptively changing at each iteration of calculations, t. Initially the weights are randomly assigned in small values. When the input vector is sent through the network, each neuron of the network computes the summed Euclidean distance between the weight and input. Among all $M \times N$ neurons, the best matching unit (BMU) which has the minimum distance between weight and input vectors becomes the winner. The BMU and its neighborhood neurons are allowed to learn by changing the weights in the manner to further reduce the distance between the weight and the input vector as shown below:

$$w_{ijk}(t+1) = w_{ijk}(t) + \eta(t)(x_i - w_{ijk}(t)) \qquad (3)$$

where $\eta(t)$ (e.g. 0.1 - 0.4) denotes the fractional increment of the correction. The radius defining neighborhood is usually set to a larger value early in the training process, and is gradually reduced as convergence is reached. The detailed algorithm of SOM can be found in [21] and [30] for ecological applications.

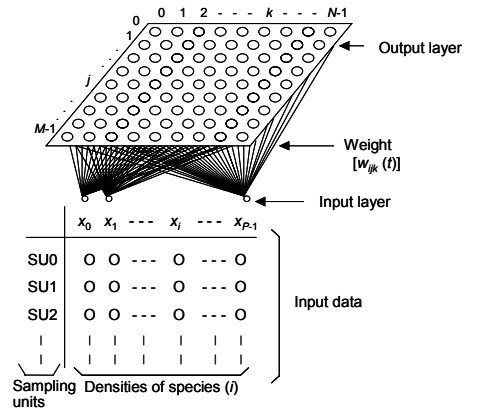

Fig. 3. Schematic diagram of SOM.

After training, the Ward's linkage method based on the Euclidian distance was additionally applied to the weights of the nodes in SOM for further clustering [31]. After preliminary training, we selected 30 ($M \times N = 6 \times 5$) units for the best output of SOM in this study. For training and visualization, we used the functions provided in the SOM toolbox [32] in Matlab. As stated before we used P=18 and P=16 species for training microorganism and macroinvertebrate communities respectively, while P=34 species were used for integrative training.

To test the null hypotheses of no significant differences in the data for environment in different clusters, we carried out nonparametric multiple comparisons after the Kruskall-Wallis test with the unequal number of samples [33].

3 Results

3.1 Patterning of microorganisms

Grouping of the sample sites was accordingly formed responding to the impact of pollution, when community data for microorganisms were trained with SOM (Fig. 4). A majority of the samples sites for the clean area (DUK) were separately grouped in the bottom right corner of the map (cluster A), while most of the samples sites from the polluted site (DKS) were placed in the top right corner of the map. According to clustering based on the Ward's linkage method, vertical clustering was formed, while horizontal subclustering subsequently occurred. The vertical clustering was observed according to the levels of pollution. While a majority of samples from the clean (DUK) and recovering (DAG) sites were observed at the bottom area (clusters A and B) of the map, the samples from the polluted sites (DKS, DDK and DKH) were located at the top area of the map (Fig. 4). In comparison with clustering, subclustering was more specific to the sites. The majority of sample site from DKS belonged to cluster D and was subsequently separated from the groups of the other polluted sites, DKH and DDK in cluster C. The sites DDK and DKH were located together in between two pollution sources (Fig. 1), and were closer each other regarding the impact of pollution compared with DKS. Similarly a large number of sample sites from the clean site DUK were located in cluster A and was separated from the groups including the recovering site DAG.

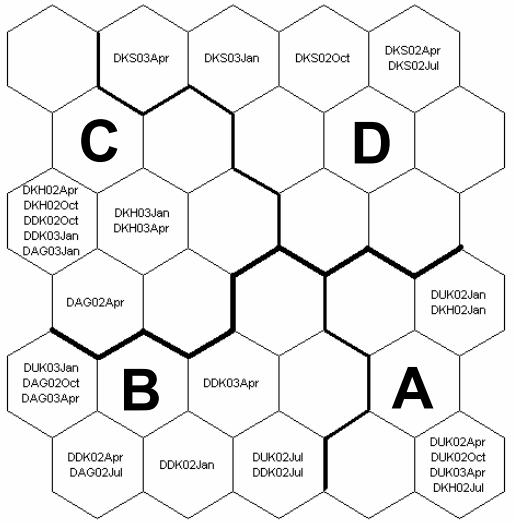

Fig. 4. Clustering of the sample sites on SOM based on the community data of microorganisms collected in the Daechon stream across different levels of pollution. Acronyms in units on the map stand for the samples: the first three alphabets represent the sites, while the following numbers and three alphabets indicate year and month of collection (e.g. DUK02Jan; site DUK collected in January 2002). (Unit of the Ward's linkage method: border for clustering; 2.5, and border for subclustering; 1.0)

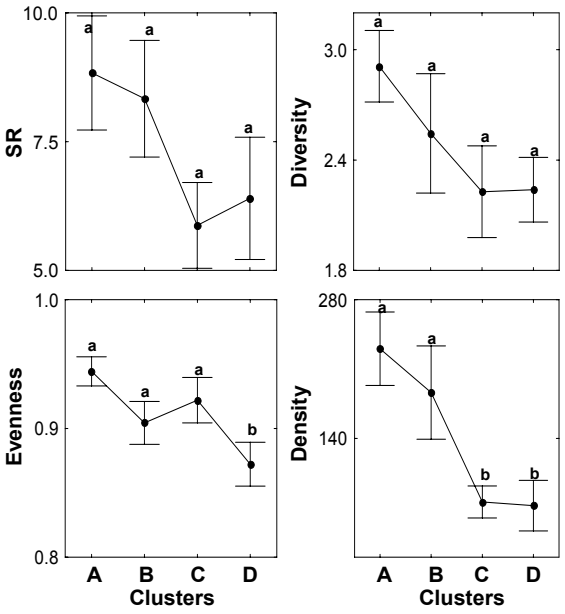

Fig. 5. Comparison of community indices in different clusters on SOM based on microorganisms (Fig. 4). The same alphabets indicate no significant difference among different clusters according to the Kruskall-Wallis test.

Community structure was further analyzed in different clusters on SOM (Fig. 5). Community indices such as SR, Diversity and Evenness were not significantly different among different clusters. Density was only significantly different between clusters A-B and clusters C-D (Fig. 5).

Table 1. Variation in environmental factors with averages (min.-max.) at different sample sites clustered by SOM based on microorganisms (Fig. 4).

Environmental Factors \ Clusters (n)	A (6)	B (9)	C (8)	D (5)
BOD (*mg/L*)	12.79 (1.73-43.30)	20.54 (5.07-42.70)	39.68 (15.33-88.20)	30.55 (9.33-47.40)
Conductivity (㎲)	90.36 (35.20-245.00)	74.14 (5.8-110.40)	80.00 (21.10-133.50)	105.22 (35.40-135.40)
DOC (*mg/L*)	2.65 (1.53-6.84)	2.72 (1.13-4.94)	3.40 (1.20-6.76)	2.50 (1.80-4.12)
Ammonium-Water (㎛)	42.21 (0.00-211.20)	17.52 (0.30-126.11)	39.76 (0.00-153.08)	73.49 (0.64-205.00)
Nitrate-Water (㎛)	86.30 (43.87-114.80)	86.32 (42.12-142.08)	114.81 (89.31-147.32)	104.47 (86.42-145.94)
Phosphate-Water (㎛)*	1.70 [b] (0.05-8.14)	1.47 [b] (0.12-3.26)	1.62 [b] (0.42-3.52)	4.61 [a] (2.09-8.35)
Ammonium-Sediment (㎛)	75.51 (4.75-222.89)	191.48 (7.72-510.33)	256.49 (41.83-583.89)	120.11 (33.59-207.12)
Nitrate-Sediment (㎛)	70.05 (36.08-165.68)	35.34 (4.09-77.96)	30.83 (2.27-99.74)	86.14 (5.50-187.44)
Phosphate-Sediment (㎛)	26.30 (0.20-148.42)	3.45 (0.31-7.09)	3.99 (0.54-11.03)	1.54 (0.80-2.72)

The same alphabets indicate no significant difference among different clusters based on the Kruskall-Wallis test. * $p<0.05$

When environmental factors were compared among different clusters patterned on SOM based on microorganism communities (Fig. 4), only phosphate in water was different among the clusters (Table 1). Phosphate in water was high in cluster D (4.61 ㎛ value), compared with other clusters (1.70 ㎛, 1.47 ㎛ and 1.62 ㎛, respectively for clusters A, B and C). Considering Fig 4 and Table 1 together, patterns of microorganism communities did not clearly represent differences in environmental impact, although community grouping was observed on SOM (Fig. 4).

3.2 Patterning of macroinvertebrates

Subsequently, the sample sites based on community data for benthic macroinvertebrates were trained with SOM (Fig. 6). Grouping of the samples sites appeared more clearly according to the impact of pollution compared with grouping based on microorganisms (Fig. 4). The Ward's linkage method revealed clustering and subclustering. Firstly the samples were divided in two groups in vertical positions, polluted site in the upper area, and clean and recovering site in the lower area. The samples of polluted sites were collectively located in a large group in cluster 4 in the upper area of the map. The samples from the clean

(DUK) and recovering (DAG) sites were located in clusters 1 and 2 in the lower areas. Seasonal grouping was revealed in subclusters. In the lower clean area, cluster 1 was mainly occupied by the samples of DUK and DAG in April, while the samples collected from DUK and DAG in January and October were placed in cluster 2. In the upper area, the samples collected during the flooding season in July were mainly grouped in cluster 3. Community grouping in cluster 3 confirmed the effect of flooding in July due to the typhoon frequently occurring in the Far East Asia. Compared with grouping by microorganisms (Fig. 4), grouping was more clearly produced responding to the impact of anthropogenic (i.e. pollution) and natural (i.e. flooding) disturbances on SOM based on macroinvertebrates. This community grouping indicated that benthic macroinvertebrate would serve as an indicator group in revealing external environmental effects in aquatic ecosystems.

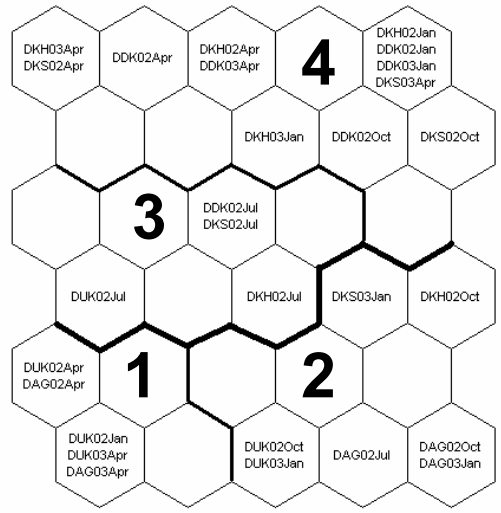

Fig. 6. Clustering of the sample sites on SOM based on the community data of macroinvertebrates collected in the Daechon stream across different levels of pollution. The acronyms are explained in Fig. 4. (Unit of the Ward's linkage method: border for clustering; 2.5, and border for subclustering; 1.2)

Differences in community indices were correspondingly revealed in different clusters on the SOM based on macroinvertebrates (Fig. 7). In contrast to the case of microorganisms (Fig. 5), SR, Diversity and Evenness were significantly different among different clusters. SR and Diversity were lower in clusters 3 and 4 representing polluted sites. In both indices, the maximum values were shown in cluster 1, while the minimum values were observed in cluster 4. This change in community indices confirmed the previous reports that SR and Diversity indices decrease as pollution levels increase (e.g. [1]). Evenness also decreased as pollution levels increased. In contrast to microorganisms Density was not significantly different among different clusters in case of macroinvertebrates.

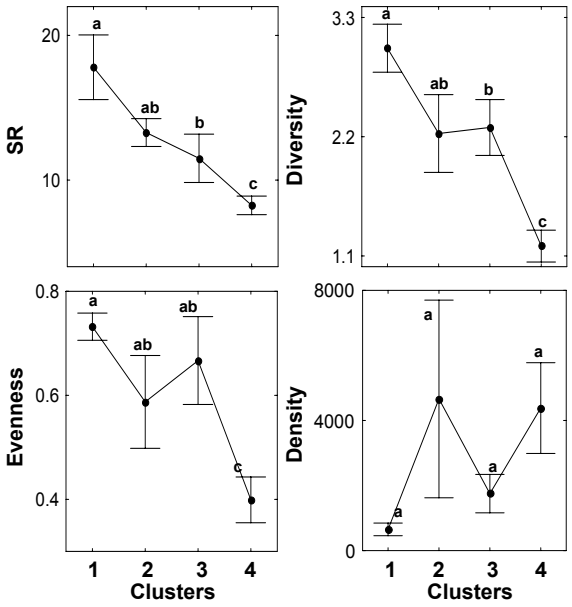

Fig. 7. Comparison of community indices in different clusters on SOM based on macroinvertebrates (Fig. 6). The same alphabets indicate no significant difference among different clusters according to the Kruskall-Wallis test.

Environmental factors based on benthic macroinvertebrates were correspondingly more discernable among different clusters (Table 2) compared with the case of microorganisms (Table 1). The grouping was clearly differentiated at the higher level of clustering between clusters 1 – 2 and clusters 3 – 4 (Fig. 6). A majority of environmental variables such as conductivity, ammonium and phosphate in water, and DOC were significantly higher in clusters 3 - 4 compared with clusters 1 - 2. Differences in subclusters were also observed. Between clusters 3 and 4 in the polluted sites, DOC was statistically different. BOD also tended to be gradually different among different clusters.

It was notable that benthic macroinvertebrates represented the gradient of pollution accordingly and additionally revealed the effect of natural disturbances, flooding also (cluster 3 in Fig. 6). Although the overall grouping was not clearly observed, however,

microorganisms also revealed additional aspect. Community was able to show local differences in the sample sites (Fig. 4). For instance, the strongly polluted site (DKS) was separated from the other polluted sites DDK and DKS which were located separately from DKS in between the first and second sources of organic pollution as stated before (Fig. 1). Similarly, the samples from the clean and recovering sites were more separated on SOM (e.g. DUK in cluster A in Fig. 4) compared with grouping by macroinvertebrates (Fig. 6).

Table 2. Variation in environmental factors with averages (min.-max.) at different sample sites clustered by SOM based on microorganisms (Fig. 6).

Environmental factors \ Clusters (n)	1 (5)	2 (7)	3 (4)	4 (12)
BOD (*mg/L*)*	7.10^{c} (1.73-20.00)	21.35^{bc} (2.13-36.00)	41.10^{a} (37.30-43.30)	33.75^{ab} (10.00-88.20)
Conductivity (μs)*	59.65^{b} (35.90-96.60)	44.56^{b} (5.80-89.00)	107.00^{a} (96.00-114.60)	113.92^{a} (32.30-245.00)
DOC (*mg/L*)*	2.02^{b} (1.53-2.99)	1.88^{b} (1.13-2.88)	2.16^{b} (1.96-2.30)	3.86^{a} (1.80-6.84)
Ammonium-Water (μm)**	0.25^{b} (0.00-0.67)	1.12^{b} (0.01-3.67)	57.85^{a} (21.68-110.38)	56.48^{a} (0.67-211.20)
Nitrate-Water (μm)	86.68 (73.18-100.41)	91.71 (42.12-142.08)	100.44 (88.69-114.80)	104.84 (77.50-147.32)
Phosphate-Water (μm)*	0.49^{b} (0.05-1.21)	1.15^{b} (0.09-2.51)	2.79^{a} (1.65-4.84)	2.69^{a} (0.40-8.14)
Ammonium-Sediment (μm)	67.30 (4.75-257.47)	210.15 (49.57-583.89)	134.93 (7.72-318.69)	206.80 (11.27-510.33)
Nitrate-Sediment (μm)	59.99 (27.35-106.05)	43.70 (5.50-77.96)	98.69 (12.50-165.68)	37.67 (2.27-187.44)
Phosphate-Sediment (μm)	1.45 (0.20-2.78)	4.52 (0.31-11.03)	2.84 (1.18-6.80)	14.85 (0.54-148.42)

The same alphabets indicate no significant difference among different clusters based on the Kruskall-Wallis test. * p<0.05, ** p<0.01

The separate analyses on both microorganisms and macroinvertebrates showed that two communities differently revealed ecological status of polluted ecosystems. While macroinvertebrates indicated overall status of water quality covering different sources of disturbances, macroorganims were more specific in representing differences in the local sample sites. Compensation between two communities may be conveyed on the integrative analysis. Subseqeutnly we carried out community patterning on the merged data for two taxa.

3.3 Integrative analysis

When the combined data for abundance of microorganism and macroinvertebrate communities were trained with SOM, the community patterns observed in Figs. 4 and 6 were revealed in an integrative manner on SOM (Fig. 8). Clustering and subclustering were similarly observed. Firstly, the patterned communities were vertically divided: clusters I-II for the samples from the clean sites in the upper area of the map, and clusters III-IV for the samples from polluted sites in the bottom area (Fig. 8).

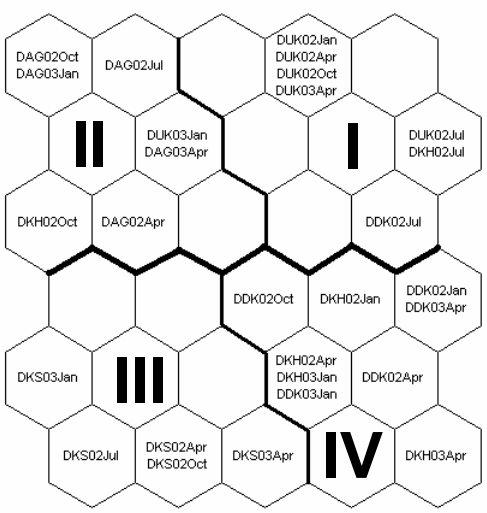

Fig. 8. Clustering of the sample sites on SOM based on the combined community data (relative abundance) of microorganisms and macroinvertebrates collected in the Daechon stream across different levels of pollution. The acronyms are explained in Fig. 4. (Unit of the Ward's linkage method: border for clustering; 3.0, and border for subclustering; 1.5)

Sub-clustering was further revealed within the groups. In the clean area, the sample sites from DUK in the upstream area in cluster I was subdivided from the samples from the recovering site, DAG, in cluster II. In clusters III-IV representing polluted sites, the samples from DKS in cluster III located below the second source of pollution was separately grouped from the sample sites located in between the first and second sources of organic pollution, DKH and DDK, in cluster IV (Fig. 8). This type of subclustering appeared to be influenced by grouping based on microorganisms (Fig. 4). The samples influenced by flooding were observed also on Fig. 8, somewhat smaller, but as a group of sample sites within cluster 1. The group of the sample sites in July, however, was placed within the clean zone in cluster I in integrative SOM (Fig. 8), while the group was located within the polluted zone on SOM based on macroinvertebrates, although the group was located close to the clean zone (Fig. 6). The shift from the polluted zone to the clean zone was influenced by groupings based on microorganisms (Fig. 4). The samples belonged to the group in July in Fig. 8 were all placed in clusters B and A in the clean zone in Fig. 4 (i.e. DUK02Jul,

DDK02Jul and DKH02Jul). It is notable that communities affected by flooding were differently evaluated in different taxa, and further investigation is needed in this regard in the future.

The spectrum of macroinvertebrates and microorganisms were also comparable according to the map. Densities of the selected taxa in averages appeared differently (Fig. 9) corresponding to the sample sites clustered by Fig. 8. In both macroinvertebrates and microorganisms, there were taxa broadly covering the map. In macroinvertebrates, for instance, Gomphidae was widely present, being concentrated more at the upper area of map around clusters I and II. Chironomidae and Oligochaeta were abundant at the lower area with higher densities concentrated at cluster IV. In microorganisms, *Exiguo bacterium* occurred in a broad scope (Fig. 9). Unidentified *beta proteobacterium* and *Bacillus* also covered a broad area. However, the widely distributed species in microorganisms were not as abundant as shown in macroinvertebrates. Some taxa were selectively present in the limited area (Fig. 9). Species in macroinvertebrates appeared more selectively at the clean area (e.g. Lepidostomatidae (cluster II) and Leuctridae and Perlidae (cluster I)), while presence of the selective species in microorganisms were more limited to the polluted sample sites (e.g. *Acidovorax* and *Nostocoida* in cluster III).

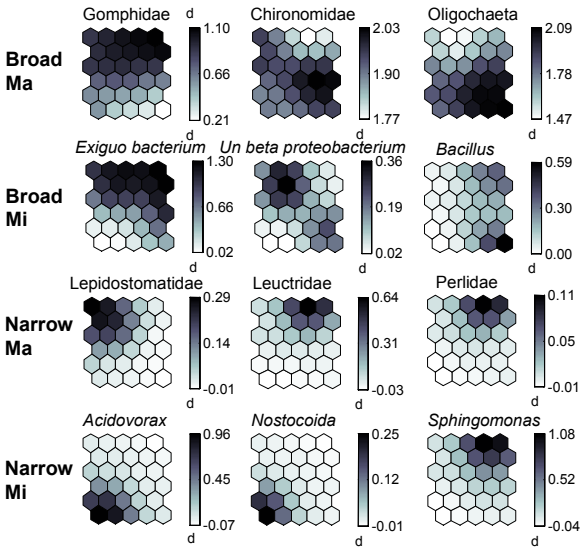

Fig. 9. Densities in averages for the taxa correspondingly appearing at the sample sites patterned by SOM (Fig. 8). Darker color represents higher densities, while the numbers along the vertical boxes indicate values of densities. (Ma: Macroinvertebrates, Mi: Microorganisms)

Table 3 summarizes the taxa appearing in broad and narrow scopes on the sample sites clustered by SOM (Fig. 8). Species in macroinvertebrates were more broadly abundant at clean sites (clusters I and II), while species in microorganisms diversely appeared at polluted sites. Additionally, more species in microorganisms appeared selectively in the polluted area, while the selective species in macroinvertebrates were more present in the clean and recovering area. In macroinvertebrates, no selective species was observed at the polluted site, DKS.

Table 3. Abundant taxa collected at the sample sites in different clusters. (Bold characters indicate the taxa selectively appearing in each cluster.)

Cluster (Site)	Macroinvertebrate	Microorganism
I (DUK)	**Ephemeridae, Leuctridae, Perlidae,** Tipulidae, Heptageniidae, Leptophlebiidae, Gomphidae, Baetidae, Gammaridae	***Janthinobacterium, Sphingomonas,*** *Exiguobacterium* *Bacillus*
II (DAG)	**Pleuroceridae, Ephemerellidae, Hydropsychidae, Rhyacophilidae, Lepidostomatidae,** Tipulidae, Heptageniidae, Leptophlebiidae, Gomphidae, Baetidae, Gammaridae, Chironomidae	**Unidentified *beta proteobacterium,*** *Exiguobacterium*
III (DKS)	Baetidae, Chironomidae, Hirudinae, Oligochaeta	***Acidovorax, Aeromonas, K. pneumoniae, Klebsiella, Luteimonas, Nostocoida, Stenotrophomonas,*** **Unidentified *Cytophagales***
IV (DKH, DDK)	**Planariidae,** Gammaridae, Chironomidae, Hirudinae, Oligochaeta	***Acinetobacter, Weissella cibaria, Weissella soli, Bacillus,*** *Exiguobacterium*

Community indices were compared in different clusters patterned by the integrative SOM (Fig. 10). Similar to comparisons in community indices in microorganisms (Fig. 5) and macroinvertebrates (Fig. 7), significant differences in SR, Diversity and Evenness were observed between clusters in clean and polluted zones in case of macroinvertebrates (Fig. 10). Differences in indices were more discernable between clusters I-II and clusters III-IV. Similar to Fig. 7, however, Density was not different among clusters. SR and Diversity were not significantly different among different clusters in communities of microorganisms as similarly shown in Fig. 5. However, Evenness was different among clusters, while Density was not different in different clusters (Fig. 10) based on the integrative SOM (Fig. 8) in contrast to the results shown in Fig. 5. The reasons for these changes in differences in Evenness and Density on the integrative SOM, however, are presently unknown. More detailed study on relationships between two

communities in habitats needs to be carried in relation to environmental effects in the future.

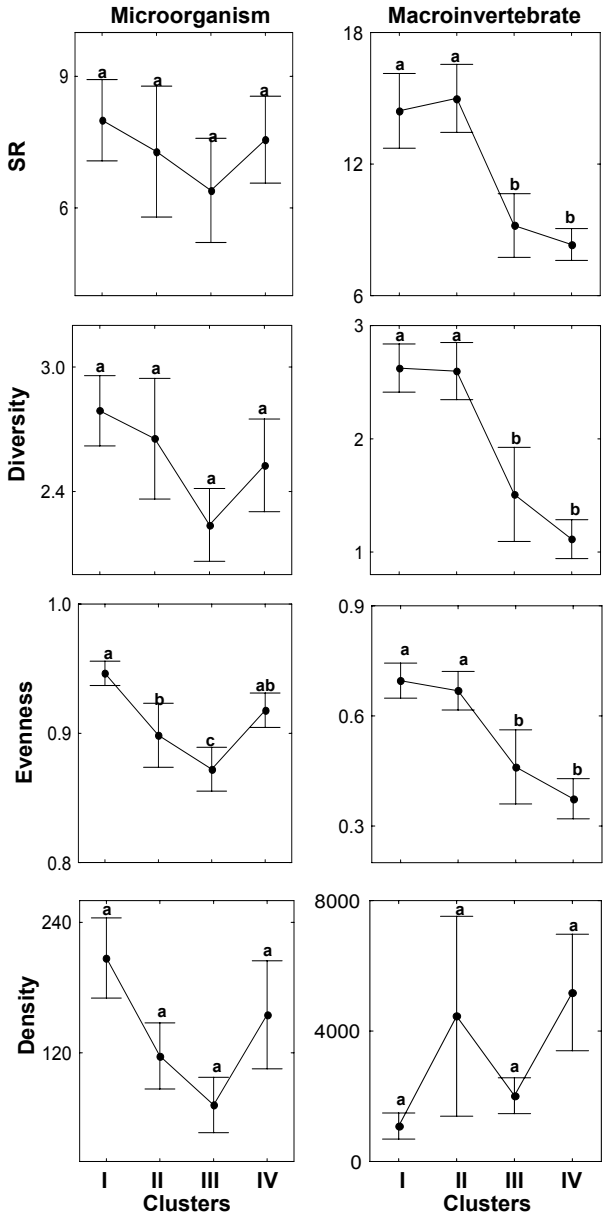

Fig. 10. Comparison of community indices in different clusters on SOM based on combined data of microorganisms and macroinvertebrates (Fig. 8). The same alphabets indicate no significant difference among different clusters according to the Kruskall-Wallis test.

Environmental variables were correspondingly different in the different sample sites when the variables were arranged according to the clusters defined by SOM (Fig. 8). The differences among clusters were more clearly addressed in the integrative patterning than the separate patterning on microorganisms (Fig. 4) and macroinvertebrates (Fig. 6). Most environmental variables showed low values in cluster I (DUK site) in the clean area (Fig. 11). In contrast, the sites DKH and DDK in cluster IV in the polluted area showed the highest levels in environmental variables in most cases.

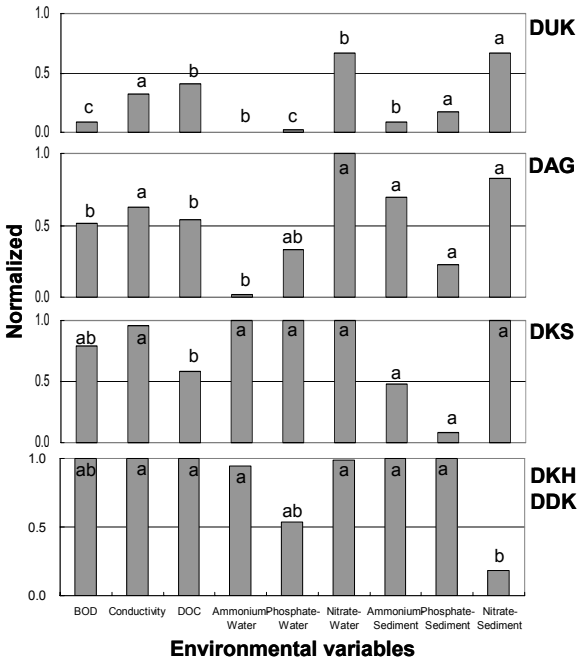

Fig. 11. Variation in environmental factors in different sample sites clustered by SOM based on the integrative data of microorganisms and macroinvertebrates (Fig. 8). The same alphabets on the vertical bar indicate no significant difference among different sample sites based on the Kruskall-Wallis test.

While the patterns were similar for the most variables between the clean and recovering sites, phosphate in water, nitrate in water, and ammonium in sediment were higher in the recovering site DAG than in the clean site DUK. At the polluted site DKS (cluster III), DOC was lower while nitrate in sediment was higher compared with DKH and DDK (cluster IV). Although the levels of phosphate in sediment appeared to be different among the sample sites, the statistical analysis based on the Kruskall-Wallis test did not show significant difference. This was due to the fact that extremely high values were obtained in DKH (148.4 μm) in cluster IV, while the other values were small (< 11.0 μm) in a great degree. A substantial difference in conductivity was also observed with no statistical difference among the sample sites. In this

case, however, *p* for alpha value was a little over 0.08, indicating that the tendency of differences among the sample sites exists. Some similar cases were also observed in the comparison of environmental factors clustered on SOM based on separate taxa (e.g. Ammonium, Phosphate and Nitrate in sediment (Tables 1 and 2)). Further study is required regarding variance in measurements in environmental factors.

Considering Figs 8-11 and Table 3, together the results suggested the relationships between environmental factors and the taxa. Ephemeridae, Leuctridae and Perlidae in macroinvertebrates, and *Janthinobacterium* and *Sphingomonas* in microorganisms (Table 3), for instance, were associated with the clean site DUK. Nitrate and phosphate in water, and ammonium in sediment at this site were significantly different from the other sites (Fig. 8). At the strongly polluted sites, DKH and DDK, Planariidae in macroinvertebrates and various species including *Acinetobacter* and *Weissella cibaria* in microorganisms appeared (Table 3) and were associated with high levels in DOC and low levels in nitrate in sediments (Fig. 11). In recovering site DAG, ammonium in water was significantly lower than the polluted site, and species in macroinvertebrates (e.g. Pleucerellidae and Ephemerellidae) and in microorganisms (e.g. Unidentified *beta proteobacterium*) appeared to be associated with this type of environmental condition. Among the same polluted sites, differentiation of environmental factors was also observed. The sample site DKS located downstream from the second source of pollution showed lower levels of DOC and higher levels of nitrate in sediment in comparison with two polluted sites, DKH and DDK, located upstream in between the first and second pollution sources (Fig. 1). At DKS there was no selective species occurring in macroinvertebrates, while various species in microorganisms were selectively associated at this site.

The overall scopes of macroinvertebrate and microorganism communities and their associations with environmental factors were revealed through the inter-taxa data analysis with SOM. At this stage, however, it is not easy to disclose the selective appearance for each species specifically, and the topic is beyond the scope of this study. Detailed investigation on distribution of each species would be discussed elsewhere. The results from this study, however, demonstrated that the inter-taxa community analysis would be useful for characterizing overall scopes of communities by comprehensively reflecting the component communities (i.e. macroinvertebrates outlining the overall impact of disturbances and microorganisms responding to local differences in the sample sites) in different environmental conditions and that the procedure for dealing with complex data could be carried out with assistance of techniques in ecological informatics such as SOM.

We also demonstrated that the full spectrum of *in situ* bacterial communities in polluted streams could be efficiently revealed through DGGE analysis. Feasibility of DGGE in identifying microorganism species from field collections confirmed the possibility of using microorganisms as a means of field bioindicators [34].

4 Conclusions

DGGE was efficient in revealing *in situ* microbial communities from field conditions: microorganism communities could be directly compared with benthic macroinvertebrates collected from the same sample sites in field. The separate analyses on microorganisms and macroinvertebrates revealed the ecological status of polluted ecosystems in different aspects. Communities of benthic macroinvertebrates responded to overall changes in external effects including anthropogenic (i.e. pollution) and natural (i.g. flooding) disturbances, while communities of microorganisms more responded to local differences in the sample sites. The inter-taxa presentation was further efficient in revealing polluted states in a broad scope in aquatic systems. Microorganisms were more associated with the polluted sites, while macroinvertebrates were more related to the clean and recovery sites. In dealing with the complex inter-taxa community data, SOM was efficient in clustering and visualization of community association, and showing relationships with environmental variables. The multi-taxa presentation would be an efficient means for revealing the full scope of polluted state of aquatic ecosystems to achieve integrative assessment of aquatic ecosystems with assistance of the techniques in ecological informatics.

Acknowledgment:
This work was supported by the Korea Research Foundation Grant (KRF-2002-015- CP0413).

References:

[1] Hellawell J.M., *Biological Indicators of Freshwater Pollution and Environmental Management*. Elsevier, London, 1986.

[2] Rosenberg D.M., Resh V.H., *Freshwater Biomonitoring in Benthic Macroinvertebrates*. Chapman and Hall, New York, 1993.

[3] Wright J.F., Sutcliffe D.W. Furse M.T., *Assessing the biological quality of fresh waters: RIVPACS and other techniques.* Freshwater Biological Association, Ambleside, UK, 2000.

[4] Butcher J.T., Stewart P.M., Simon T.P., A benthic community index for streams in the northern lakes and forests ecoregion. *Ecological indicators*, 3, 2003, 181-193.

[5] Monster U., Chrost R.J., Orgin, composition and microbial utilization of dissolved organic matter. 00. 8-46. In Overbeck, J., Chrost R.J. (eds). *Aquatic microbial ecology; biochemical and molecular approaches.* Springer Verlag, New York, 1990.

[6] Allan, J.D., *Stream Ecology 'Structure and function of running waters'.* Chapman & hall, 1995.

[7] Porter K.G., Feig Y.S., The use of DAPI ofr identifying and counting aquatic microflora. *Limnol. Oceanogr*. 25, 1980, 943-948.

[8] Pernthaler J., Glockner F-O., Unterholzner S., Alfreider A., Psenner R., Amann R., Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. *Appl. Environ. Microbiol.,* 64, 1998, 4299-4306.

[9] Cody D.G., Heath R.T., Leff L.G., Characterization of benthic bacterial assemblages in a polluted stream using denaturing gradient gel electrophoresis. *Hydrobiologia*, 432, 2000, 207-215.

[10] Bonin H.L., Griffiths R.P., Caldwell B.A., Nutrient and microbiological characteristics of fine benthic organic matter in mountain streams. *J. N. Am. Benthol. Soc*., 19, 2000, 235-249.

[11] Richards C., Haro R.J., Johnson L.B., Host G.E., Catchment and reach-scale properties as indicators of macroinvertebrate species traits. *Freshwater Biology*, 37, 1997, 219-230.

[12] Yeager P.E., Foreman C.L., Sinsabaugh R.L., Microbial community structure and function in response to larval chironomid feeding pressure in a microcosm experiment. *Hydrobiologia*, 448, 2001, 71-81.

[13] Leff L.G., Lemke M.J., Ecology of aquatic bacterial populations: lessons from applied microbiology. *J. N. Am. Benthol. Soc*., 17, 1998, 261-271.

[14] McEwen, H.A., Leff L.G., Colonization of stream macroinvertebrates by bacteria. *Arch. Hydro.* 151, 2001, 51-65.

[15] Lemly A.D., Bacterial growth on stream insects: potential for use in bioassessment. *J. N. Am. Benthol. Soc*., 17, 1998, 228-238.

[16] Hieber M., Gessner M.O., Contribution of stream detrivores, fungi, and bacteria to leaf breakdown based on biomass estimates. *Ecology*, 83, 2002, 1026-1038.

[17] Haapala A., Muotka T., Breakdown and macroinvertebrate and fungal colonization of alder, birch, and willow leaves in a boreal forest stream. *J. N. Am. Benthol. Soc*., 20, 2001, 395-407.

[18] Graca M.A.S., Ferreira R.C.F., Coimbra C.N., Litter processing along a stream gradient: the role of invertebrates and decomposers. *J. N. Am. Benthol. Soc*., 20, 2001, 408-420.

[19] Carlisle D., Clements W.H., Leaf litter breakdown, microbial respiration and shredder production in metal-polluted streams. Freshwater biology, 50, 2005, 380-390.

[20] Kohonen T., *Self-organizing maps*. 3rd ed. Springer, Berlin, 2001.

[21] Chon, T.-S., Park, Y.S., Moon, K.H., Cha, E.Y., Patternizing communities by using an artificial neural network. *Ecol. Model*., 90, 1996, 69-78.

[22] Park, Y-S, Chon, T-S., Kwak, IS., Lek, S., Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. *Science of the Total Environment,* 327, 2004, 105-122.

[23] Urakawa, H., Noble P.A., Fantroussi S.E., Kelly J.J., Stahl D.A., Single-base-pair discrimination of terminal mismatches by sing oligonucleotide microarrays and neural network analyses. *Appl. Environ. Microbiol.*, 68, 2002, 235-244.

[24] Wang, H-C., Badger J., Kearney P., Li M., Analysis of codon usage patterns of bacterial genomes using the self-organizing map. *Mol. Biol. Evol*., 18, 2001, 792-800.

[25] APHA, AWWA, WPCF, *Standard methods for the examination of water and waste* (16th ed.). Washington D.C., 1985.

[26] Muyzer, G, Waal EC, Uitterlinden AG, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl and Envir. Microbiol.,* 59, 1993, 695-700.

[27] Merritt R.W., Cummins K.W., *An Introduction to the Aquatic Insects of North America*. Hunt Publishing Company, Dubugue, 1984.
[28] Pennak R. W. 1978. - *Fresh-water Invertebrates of the United States*. John Wieley & Sons, Inc., New York. 803 p.
[29] Brinkhust R. O. 1986. - *Guide to the Freshwater Aquatic Microdrile Oligachaetes of North America*. Canadian Special Publication of Fidheries and Aquatic Sciences 84, 259 p.
[30] Giraudel JL, Aurelle D., Berrebi P., Lek S., Application of the self-organising mapping and fuzzy clustering to microsatellite data: how to detect genetic structure in brown trout (Salmo trutta) populations. Pages 187-202 in *Artificial Neuronal Networks: Application to Ecology and Evolution.* Lek S., Guégan J-F. (eds). Springer-Verlag, Berlin, 2000.
[31] Jain A.K., Dubes R.C., *Algorithms for clustering data*. Prentice-Hall, Englewood Hills, NJ. 1988.
[32] Alhoniemi E., Himberg J., Parhankangas J., Vesanto J., SOM Toolbox. [online] http://www.cis.hut.fi/projects/somtoolbox, 2000.
[33] Zar J.H., *Biostatistical Analysis*. 4th ed.. Prentice-Hall Inc., New Jersey, 1999.
[34] Van Beelen P., Doelman P., Significance and application of microbial toxicity tests in assessing ecotoxicological risks of contaminants in soil and sediment*, Chemosphere*, 34, 1997, 455-499.