Korean J. Limnol. 42 (3) : 303~316 (2009)

Comparative Analyses of Community and Biological Indices based on Benthic Macroinvertebrates in Streams using a Self-Organizing Map

Tang, Hong Qu, Mi-Jung Bae, Tae-Soo Chon[1], Mi-Young Song[2] and Young-Seuk Park*

Department of Biology and The Korea Institute of Ornithology, Kyung Hee University, Seoul 130-701, Korea
[1]Department of Biological Sciences, Pusan National University, Busan (Pusan) 609-735, Korea
[2]West Sea Fisheries Research institute, Incheon 400-420, Korea

Benthic macroinvertebrate communities collected from eight different streams in South Korea were analyzed to compare community and biological indices across different levels of water pollution. The Self-Organizing Map (SOM) was utilized to provide overview on association of the proposed indices. The sample sites were accordingly clustered according to the gradient of pollution on the SOM. While the general trends of the indices were commonly observable according to different levels of pollution, the detailed differences among the indices were also illustrated on the SOM. The conventional diversity and evenness indices tended to be high even though the water quality state was poor representing relatively weak gradient at polluted sites, while the index presenting the saprobic degree such as family biotic index showed the stronger gradient at the polluted area and was robust to present the gradient. Our results also confirmed the general characterization of two indices: The Shannon index is more strengthened by the number of species occurring at the sample sites, while the Simpson index is more influenced by the degree of evenness among the species. The patterning based on the SOM was efficient in comparatively characterizing the proposed indices to present ecological states and water quality.

Key words : community indices, biological indicators, diversity, water quality, Self-Organized Map, pollution

INTRODUCTION

Benthic macroinvertebrates have been studied widely and have been considered as one of the most suitable indicator groups for assessing aquatic ecosystems (Hynes 1960; Hellawell, 1986; Rosenberg and Resh, 1993; Barbour *et al.*, 1996; Butcher *et al.*, 2003; Park *et al.*, 2007). Benthic macroinvertebrates are sedentary and have optimally intermediate life span (from months to a few years). They, therefore, are suitable for presenting water quality in collective and integrative manner. In addition, they play a key role in food web dynamics, linking producers and top carnivores (Rosenberg and Resh, 1993; Song *et al.*, 2007).

Since Kolkwitz and Marsson pioneered in measuring saprobien system in lakes in 1902, a number of water quality indices in early stage have been proposed in 1970's in aquatic ecosystems (Wihlm, 1972; Hawkes, 1979; Sladecek, 1979), reporting Saprobien index (Sladecek, 1969, 1979;

*Corresponding author: Tel: 02) 961-0946, Fax: 02) 961-0244, E-mail: parkys@khu.ac.kr

Hawkes, 1979), Trent Biotic Index (TBI) (Woodiwiss, 1978), species diversity index (Pielou, 1966; Wilhm, 1968) and scoring system (Chandler, 1970). James and Evison (1979) discussed importance of benthic macroinvertebrates in presenting water quality.

Since 1980's a numerous account of researches have been conducted to broaden the scope of biological indices and to evaluate the proposed indices. National Water Council (1981) in UK developed a straightforward but very effective index in presenting water quality based on benthic macroinvertebrates, Biological Monitoring Working Party (BMWP) Score. From that, most biotic indices were based in score systems, such as the Biotic Condition Index (Winget and Mangun, 1979) and the Family Biotic Index (Hilsenhoff, 1987, 1988) (Rosenberg and Resh, 1993, Baptista *et al.*, 2007). Hellawell (1986) and Spellerberg (1991) evaluated performance of various biological indices and compared their advantages and disadvantages. Regarding benthic macroinvertebrates, Rosenberg and Resh (1993) provided extensive reviews on water quality indices based on benthic macroinvertebrates, focusing on rapid assessment and quantitative approaches. Recently, considering existence of numerous indicators and complexity residing in water quality, Johnson and Gordkoop (2000) checked multi-metric monitoring to present water quality and proposed the River Invertebrate Prediction and Classification System (RIVPACS) as an integrative evaluation method.

However, it is a difficult task to develop an objective and quantifiable indicator system from community data. Data for benthic macroinvertebrates are especially complex since the communities consist of multi-variables (i.e., diverse taxa) varying in a non-linear fashion. There have been numerous indicators proposed for water quality evaluation as stated above. However, not much research has been conducted to directly compare the indices to provide comprehensive information on overall associative relationships among the indices. Kerrans and Karr (1994) evaluated 18 different attributes and determined suitable (e.g., total species richness, species richness in selected taxa) and unsuitable (e.g., sediment-surface taxa richness) attributes. Recently, Gray and Delaney (2008) compared benthic macroinvertebrate indices including BMWP, ASPT (Average Score Per Taxa), and Shannon diversity index for the assessment of the impact of acid mine drainage on river, showing differences found in the ability of indices to discriminate disturbance impact. However, studies mentioned above simply discussed the indices in qualitative and descriptive manner. Direct comparison of indices has not been done under the scheme of information extraction from the multivariate data.

In this study we intended to compare the proposed indices and checked the associative relationships among indices. Considering the complexity residing in the water quality indices we applied a computational adaptive learning algorithm, a Self-Organizing Map (SOM), for mining information residing in the data of benthic macroinvertebrates in streams.

MATERIALS AND METHODS

1. Study area

Datasets were built with benthic macroinvertebrates collected at the 32 sample sites in eight streams (Suyong, Soktae, Onchon, Daechon, Hakjang, Paenae, Dobong, and Yangjae), covering two largest (the Nakdong and Han) and one shorter (the Suyong) river basins in Korea from 1992 to 2007 (Fig. 1). In total, 751 sample cases were used in the analyses. Sample sites showed various pollution levels. Based on the biotic index ASPT (Armitage *et al.*, 1983), the water quality of sample sites were classified into three categories, Clean sites (ASPT > 6): B (Paenae stream), M1 ~ M3, M5 (Dobong stream), O1 (Onchon stream), D1 (Daechon stream); intermediately polluted sites (4 < ASPT < 6): M4, M6 (Dobong stream), Y1 ~ Y3 (Suyong stream), T1, T2 (Soktae stream), O2 (Onchon stream), D2 ~ D4 (Daechon stream); polluted sites (ASPT < 4): Yangjae stream, Hakjang stream, T3 and T4 (Soktae stream) (Table 1). Clean and relatively clean sites were selected from the Paenae stream in the Nakdong River basin and the Dobong stream in the Han River basin (Fig. 1). The Paenae stream was reported to be clean and was surveyed for a long period for the Korean National LTER (long-term ecological research) (Oh and Chon, 1991a, b; Oh and Chon, 1993; Qu *et al.*, 2008) (BOD 4.19 mg L^{-1}, conductivity 23.86 $\mu S\ cm^{-1}$). They showed relatively high BMWP scores (79 ~ 140) and ASPT (7.56 ~ 8.81) (Table 1).

The Dobong stream is a tributary of the Jung-

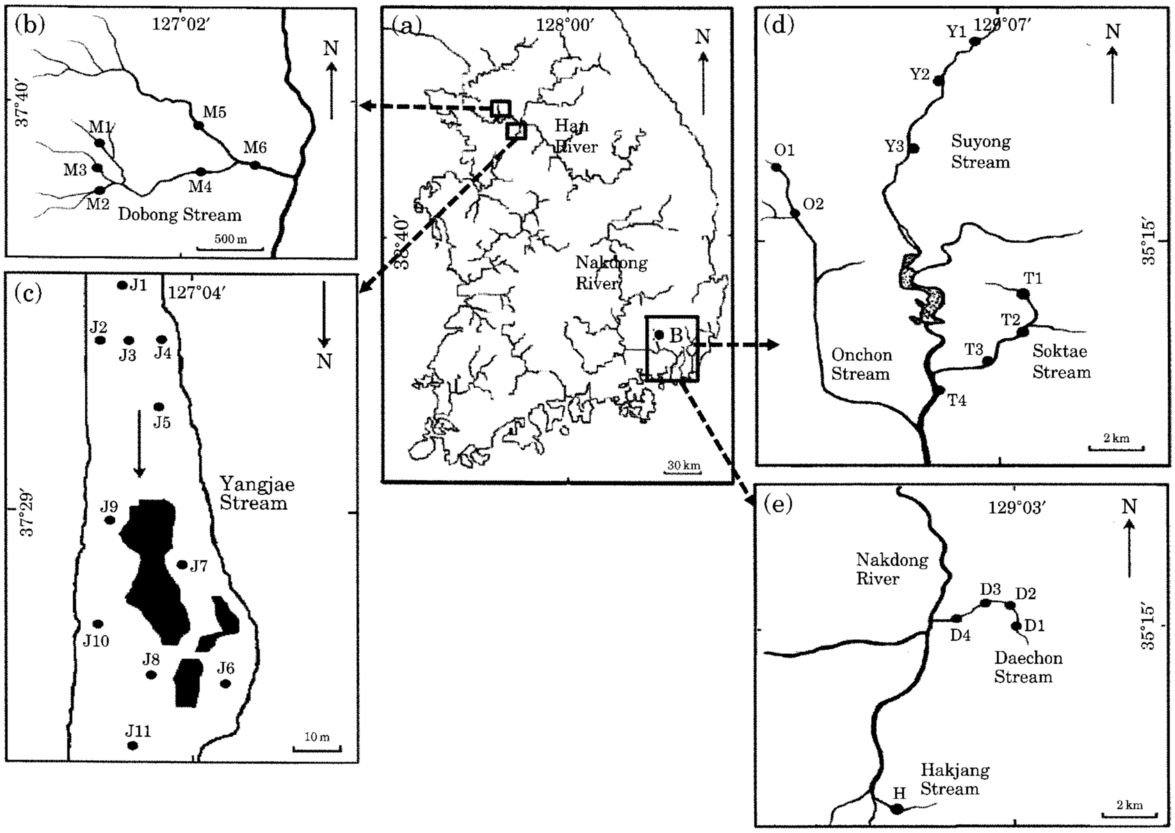

Fig. 1. Location of the sample sites in different streams in Korea. (a) Korean Peninsula showing Nakdong and Han Rivers including site B (Paenae stream), (b) Yangjae stream (J1 ~ J11) and (c) Dobong (M1 ~ M6) stream in the Han River, (d) Suyong (Y1 ~ Y3) and Soktae (T1 ~ T4) streams in the Suyong River, and (e) Daechon stream (D1 ~ D4), Onchon stream (O1, O2) and Hakjang stream (H).

rang stream, which is one of the main river systems belonging to the Han River basin. Three sites (M1 ~ M3) were selected to present forest and mountain area (conductivity 60.41 μS cm^{-1} in M1, 95.68 in M2 and 57.06 μS cm^{-1} in M3). One site (M5) was included to present influence from restoration in the mountainous urban area (conductivity 137.67 μS cm^{-1}), while another two sites (M4, M6) were located in the urban area showing relatively low BMWP scores (around 32) and ASPT (5.20 ~ 5.27) (Table 1).

The intermediate and polluted streams were also selected from the Suyong River basin in the Busan area in the south east of Korea. In this basin, three streams (Onchon, Suyong and Sokdae streams) were included. The Suyong stream suffered from intermediate pollution from agricultural practices in the watershed area and covered the sites presenting relatively clear upstream (Y1) and intermediately polluted sites in the mid stream (Y2 and Y3) (BOD 2.76 ~ 3.72 mg L^{-1}, conductivity 134.30 ~ 324.10 μS cm^{-1}) (Kwon and Chon, 1991; Park *et al.*, 2004) (Table 1). The BMWP scores ranged from 20 to 44. The Soktae stream received various sources of pollution such as agricultural, industrial and domestic sewage discharge (Kang *et al.*, 1995; Youn and Chon, 1996) and showed the longitudinal gradient of pollution from upper to down stream (sites T1 to T4) (BOD 4.87 ~ 46.21 mg L^{-1}, conductivity 192.70 ~ 618.50 μS cm^{-1}) (Park *et al.*, 2004; Qu *et al.*, 2008). The BMWP scores were around 6 in the polluted sites (T3, T4), while the scores were in the range of 19 ~ 43 in the less polluted sites (T1, T2). The Onchon stream covered clean upstream site (O1) and relatively polluted downstream site (O2). While site O1 is located in the small-scale mountain area in the city, site O2 is located in the middle of residential area and water recovery projects was conducted (Fig. 1e). The BMWP scores were distinctively dif-

Table 1. Environmental factors (mean ± SD), biological metrics, and water states measured at the sample sites.

Streams	Sites	BOD (mg L^{-1})	Turbidity (NTU)	Conductivity (μs cm^{-1})	BMWP scores	ASPT	Shannon index (H′)	Water states
Paenae	B	1.15±0.80	0.84±0.95	23.86±6.66	108 (79~140)	8.01 (7.56~8.81)	3.62 (2.49~5.01)	Clean
Dobong	M1	1.80±0.21	1.18±0.84	58.46±12.73	93 (64~116)	8.02 (7.33~9.11)	3.29 (2.31~3.75)	Clean
	M2	1.55±0.07	1.07±0.91	89.53±22.60	107 (63~144)	7.81 (7.04~8.44)	3.62 (2.41~4.21)	Clean
	M3	1.80±0.01	0.79±0.33	54.97±11.97	83 (60~128)	7.53 (6.61~8.94)	3.32 (2.86~3.87)	Clean
	M4	1.10±0.28	2.68±2.58	123.99±23.13	32 (7~47)	5.27 (3.60~6.44)	1.80 (0.66~3.04)	Intermediate
	M5	1.45±0.35	1.41±1.14	128.77±51.60	94 (46~123)	7.24 (6.47~8.06)	3.60 (2.54~4.26)	Clean
	M6	2.10±0.21	1.41±0.64	129.17±26.97	32 (7~64)	5.20 (3.58~7.06)	2.06 (1.25~2.80)	Intermediate
Suyong	Y1	3.28±3.71	2.63±2.78	134.30±96.24	32 (9~74)	5.35 (4.50~6.07)	2.50 (1.17~3.59)	Intermediate
	Y2	2.76±2.35	1.86±2.16	180.40±36.86	44 (4~75)	5.61 (3.70~7.31)	2.81 (1.27~3.69)	Intermediate
	Y3	3.72±4.46	7.51±11.70	325.10±156.47	20 (4~37)	4.98 (2.87~6.70)	2.08 (1.00~3.38)	Intermediate
Onchon	O1	2.75±1.70	2.43±1.31	72.82±43.29	123 (95~138)	7.47 (6.89~8.40)	3.79 (2.90~4.62)	Clean
	O2	3.78±2.68	7.72±10.41	237.51±147.3	39 (24~53)	4.88 (4.05~5.74)	2.00 (0.88~3.44)	Intermediate
Yangjae	J*	7.97±3.05	6.72±2.38	388.33±33.65	11 (4~33)	3.46 (2.94~4.50)	0.49 (0.00~1.38)	Polluted
Soktae	T1	4.87±4.21	11.15±8.50	192.70±73.31	43 (7~86)	5.98 (3.60~7.87)	2.73 (1.05~3.87)	Intermediate
	T2	6.94±5.61	6.12±13.24	335.40±97.61	19 (4~34)	4.27 (2.95~5.67)	2.16 (0.65~3.57)	Intermediate
	T3	43.23±17.35	25.51±15.25	571.20±151.91	6 (4~9)	3.61 (3.00~3.70)	1.13 (0.00~2.52)	Polluted
	T4	46.21±20.87	22.32±15.15	618.50±122.26	6 (4~7)	3.61 (3.50~3.70)	0.70 (0.00~2.25)	Polluted
Daechon	D1	2.02±2.34	5.54±8.54	41.36±7.66	104 (61~164)	7.21 (6.75~7.77)	3.63 (2.62~4.53)	Clean
	D2	6.88±5.68	4.16±3.97	176.76±73.21	36 (17~49)	4.71 (3.48~5.73)	2.08 (0.80~2.98)	Intermediate
	D3	4.52±3.36	2.87±2.31	211.90±97.31	43 (15~76)	4.91 (3.48~6.32)	2.24 (0.15~3.44)	Intermediate
	D4	2.56±1.63	0.96±0.78	161.03±75.16	70 (44~108)	6.48 (5.55~7.51)	3.32 (2.29~3.96)	Intermediate
Hakjang	H	31.80±23.86	12.92±9.03	480.44±189.20	19 (9~36)	4.01 (2.78~6.00)	0.52 (0.09~1.74)	Polluted

* For the fine scale in the Yangjae stream, 11 sites were treated here as one site with sharing the same environmental variables.

ferent in the range of 95~138 at O1, while 24~53 at O2 (Table 1).

In the downstream of the Nakdong River, two tributaries (Daechon and Hakjang streams) with different pollution levels were selected (Fig. 1e). The Daechon stream is an upper tributary of the Nakdong River. Two sites (D2 and D3) were chosen to check severely pollution with organic matters (i.e., domestic sewage from restaurants) (Song *et al.*, 2005), and other two sites (D1 and D4) presented clean and recovering status, respectively (BOD 2.02~6.68 mg L^{-1}, conductivity 41.36~211.90 μS cm^{-1}) (Song *et al.*, 2005; Qu *et al.*, 2008). The BMWP scores were low (36~43) at the polluted sites, while the indices were high at the clean sites (70~104). The Hakjang stream is located near the river estuary, and only one site was included in the study. The site received various pollution sources with the higher BOD (8~55) and showed the lower BMWP values (9~36).

The Yangjae stream flows through the commercial, domestic and agricultural areas in the southern part of Seoul, with relative lower diversity (0.00~1.38) and BMWP (4~33) (Kwak *et al.*, 2002; Song *et al.*, 2006). Fine scale sampling was conducted for a 200 m reach. Eleven sample sites (J1~J11) separated by 10 m increments were selected in the sampling area (Fig. 1c). A water recovery project was conducted at the study site in the Yangjae stream during the survey period (KICT, 1997, 1999).

2. Field data

Benthic macroinvertebrates were collected with a Surber net (30 × 30 cm^2, mesh size 500 μm) with three replicates at each sample site. Additionally, the physicochemical environmental variables including BOD, DO, turbidity, conductivity, velocity, and depth were measured at each sample site. The collected benthic macroinvertebrates were preserved in 70% alcohol solution. In the laboratory, the invertebrate specimens were sorted and identified mostly into species level, and counted for the number of specimens under microscopes. In the Suyong River basin, samples were

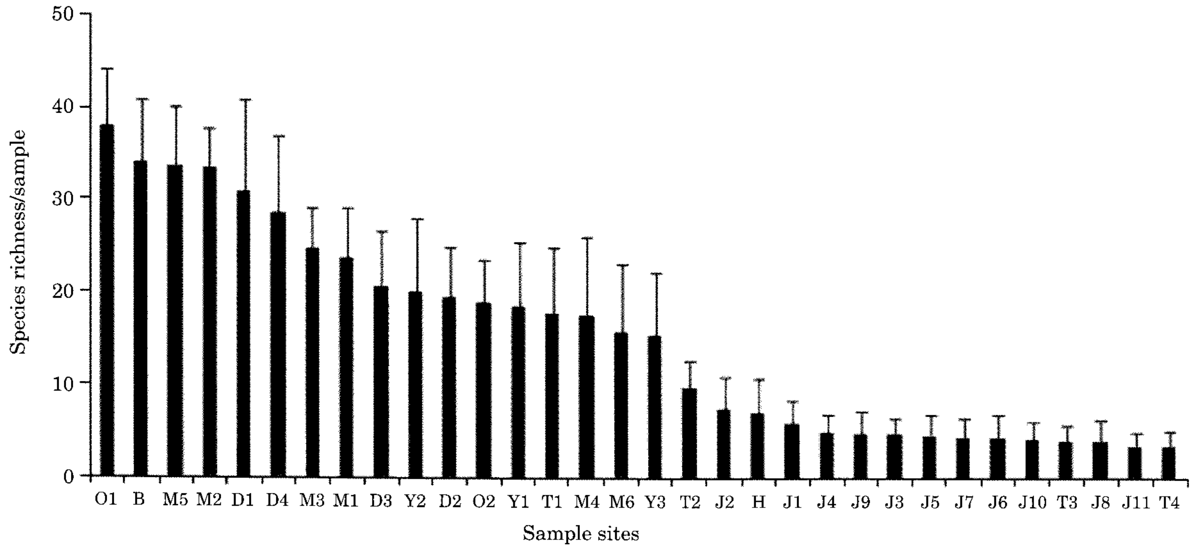

Fig. 2. The average number of species occurred in 32 sites (ranked by taxa richness). Error bars indicate standard deviations.

collected monthly for communities from March, 1992 to April 1995. In the Suyong stream, 74 samples (from three sample sites for 30 months, excluding some missing samples) were obtained with average collection of 18 species per sample, while 147 samples (from four sample sites for 48 months, excluding some missing samples) were collected and 9 species were identified in each sample in the Soktae stream (Chon and Kwon, 1991) (Fig. 2). Twelve samples (one sample site for one year) were obtained from site B in the Paenae stream from January 2007 to December 2007 with average collection of 34 species per sample, while seasonal data were conducted in the Daechon, Onchon and Hakjang streams from November 2004 to April 2007. In the Daechon stream, 88 samples (from four sample sites for 11 seasons) were obtained with identification of 25 species in average. In the Onchon stream, 22 samples (from two sample sites for 11 seasons) were collected with average 29 species per sample (Fig. 2). While in the polluted Hakjang stream, 11 samples were conducted with average seven species per sample. In the Seoul area, monthly sampling was conducted in the Dobong and Yangjae streams. In the Dobong stream, 61 samples (from six sample sites from December 2006 to November 2007, excluding some missing samples) were obtained with average collection of 26 species per sample, while 396 samples (from 11 sample sites from April 1996 to March 1999, excluding some missing samples) were collected with average five species were identified in each sample (Fig. 2).

3. Biological indices

To evaluate the scope of different biological indices and their association across different levels of pollution, we chose 18 metrics indicating biological indices based on literatures (Barbour *et al.*, 1999, Gray and Delaney, 2008). First, we included four indices related with species richness and abundance as following: total species richness (SR), total abundance (N), and EPT (Ephemeroptera, Plecoptera, and Trichoptera) species richness (EPT_SR) and its relative abundance (%EPT_N). And eight metrics were developed based on the proportions (%) of species richness and abundance of functional feeding groups (FFGs): Shredders (%SR and %N), Predators (%SR and %N), Collector-filterers (Cfilterer, %SR and %N), Collector-gatherers (Cgatherer, %SR and %N). In addition, we included three community diversity indices: Shannon index, Simpson index, and evenness. Finally, we included three biological water quality indices: Biological Monitoring Working Party (BMWP) score, BMWP average score per taxa (ASPT), and family biotic index (FBI) based on the tolerance levels of species.

4. Data analysis

We implemented the SOM (Kohonen, 1989) for checking association of indices and evaluated the scope of the indices across different levels of pollution in this study. The SOM has been extensively used for patterning community data since 1990s (Chon *et al.*, 1996, 2000, 2002; Park *et al.*, 2001, 2003a, b, 2004). The SOM efficiently mines complex data without templates (or teachers) in an unsupervised manner and has been reported as a reliable classifier of ecological data (Chon *et al.*, 1996, 2002; Park *et al.*, 2004). The modeling with SOM was conducted in two different modes according to multivariate analyses on ecological data: Q mode for sample classification and R mode for metrics classification independently.

The SOM consists of input and output layers connected with computational weights (connection intensities). The array of input neurons (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 each SOM modeling, initially the input data (18 metrics in this study) were subjected to the network. In this study, the number of output neurons were set to 140 ($=14\times10$) for sample classification and 20 ($=5\times4$) for metrics classification in 2D hexagonal lattice. A hexagonal lattice is preferred because it does not favor horizontal or vertical directions (Kohonen, 2001). The number of nodes was determined as $5\times\sqrt{\text{number of samples}}$ (Vesanto and Alhoniemi, 2000).

Subsequently, the map size was determined. Basically, the two largest eigen values of the training data were calculated and the ratio between side lengths of the map grid was set to the ratio between the two maximum eigen values. The actual side lengths were then set so that their product was close to the determined number of map units. Subsequently, the weights of the network were trained for a given dataset. Each node of the output layer computes the summed distance between weight vector and input vector. The output nodes are considered as virtual units to represent typical patterns of the input dataset assigned to their units after the learning process. Among all 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 neighborhood 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.

After training the SOM, a hierarchical cluster analysis with the Ward's linkage method based on the Euclidian distance (Ward, 1963) was applied to the weights of the nodes in the SOM for further clustering (Jain and Dubes, 1988; Park *et al.*, 2003a). For training the SOM, we used the functions provided in the SOM toolbox (Alhoniemi *et al.*, 2000) in Matlab (The Mathworks, 2001). A multi-response permutation procedure (MRPP; Mielke *et al.*, 1976), which is a non-parametric procedure for testing the hypothesis of no difference between two or more groups of entities, was conducted to evaluate the significance of the clusters. Kruskall-Wallis (K-W) test, non-paramnetric analysis of variance, was carried out to evaluate the differences of variables among clusters defined in the SOM, and then multiple comparisons of mean ranks for all groups (Siegel and Castellan, 1988) was conducted for variables in order to show the significant differences among clusters. The analyses were carried out by a statistical software STATISTICA (StatSoft, 2004).

RESULTS

Total species richness per sample at each sample site was arranged with the rank from high to low (Fig. 2). In general, species richness presented water quality of the sample sites: species richness accordingly decreased with increase in pollution levels. In order to check association among the water quality indices, 18 metrics presenting community indices and biological indicators were used as input for sample classification (Q mode) through the learning process of the SOM (Fig. 3). Based on the dendrogram of the cluster analysis with a Ward's linkage method, six clusters (1~6) were defined. The MRPP showed significant differences among clusters ($p<0.001$). The patterning of indices reflected the overall degree of disturbances pertaining to the sample sites. The samples from the clean sites were grouped in the upper areas, while the samples from the polluted sites were placed in the lower areas of the SOM map.

Most sample sites in the Yangjae and Hakjaeng streams, also including a majority of downstream sample sites (T3, T4) in the Soktae stream, belonged to this group in cluster 3 at the bottom right area of the map (Fig. 3, Table 2). At the top right corner of the map in contrast, the groups of sample sites were placed to present the clean states

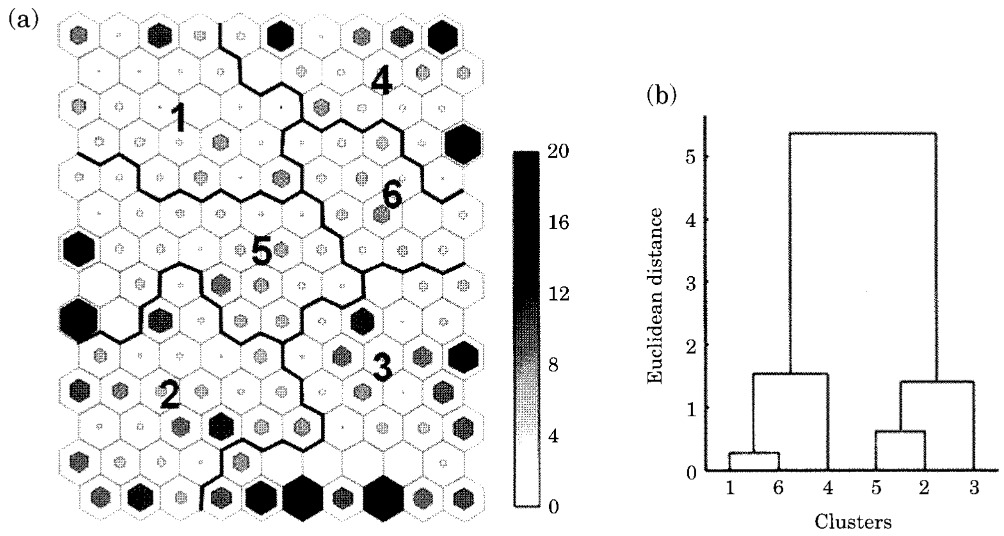

Fig. 3. (a) SOM grouping of communities based on biological indices. The different sizes of hexagon in gray scale indicate the number of samples assigned to each SOM unit as indicated by the scale bar. (b) Dendrogram of hierarchical cluster analysis with the Ward linkage method using Euclidean distance measure.

Table 2. Sample sites and the number of samples in the different clusters defined in SOM.

Clusters	Sample sites (Number of samples)												Total
1	Y1(15)	Y2(13)	Y3(6)	T1(12)	T2(26)	D2(2)	D3(2)	O2(3)	M4(2)	M6(3)	B(1)	J(2)	87
2	T3(16)	T4(12)	J(130)	H(1)									159
3	J(184)	H(8)	T4(10)	T3(6)	D2(1)	D3(3)	O2(2)	M4(1)	M6(1)				215
4	Y2(16)	T1(16)	D1(11)	D4(8)	O1(11)	M1(11)	M3(11)	M2(12)	M5(11)	B(11)	O2(2)		116
5	J(80)	T3(13)	T4(6)	T2(6)	Y3(5)	Y2(2)	D2(2)						120
6	Y3(10)	Y1(6)	Y2(5)	D2(6)	D3(6)	M4(5)	M6(2)	T1(4)	T2(3)	D4(3)	O2(3)		54

in cluster 4. Most of the sample sites of M1~M3, M5 in the Dobong stream, site B in the Paenae stream, sites D1 and D4 in the Daechon stream, and site O1 in the Onchon stream were observed in this cluster. Another large group of the sample sites were observed in the middle area belonging to cluster 5 in the left edge of the map. A part of samples from sites T3 and T4 in the Soktae stream and some sites in the Yangjae stream belonged to this cluster, presenting the polluted sites. However, the patterns of communities were somewhat different from the other polluted sites identified at the bottom right area of cluster 3 (Fig. 3).

Characterization of the grouped communities could be illustrated with visualization of input variables (Fig. 4). From visualization, the overall and specific trends in indicators could be explained. In general, the values of indicators increased along with cleanness at the area of top right corner corresponding to cluster 4 in the map. Community indices such as Shannon and Simpson diversity indices, total species richness, and biological indicators such as species richness and abundance (%) in EPT, BMWP, and ASPT showed the similar trends. Also some indicators presenting proportions of FFGs were similarly showed the highest range in cluster 4. Species richness (%) in shredders and collector-filterers also tended to show the maximal range in cluster 4. In contrast, a few indices such as FBI, species richness (%) and abundance (%) in collector-gatherers showed the higher values as the degree of pollution increased.

Although the general trends were similar, the profiles of the indicators also showed differences in detail according to the profiles of indicators (Fig. 4). While the areas of strongly higher values were narrow in some variables, the other variables were dispersed widely. Species richness and abundance (%) in EPT displayed high values in the narrow range on the SOM map, while ASPT displayed relatively wide range. Diversity indices

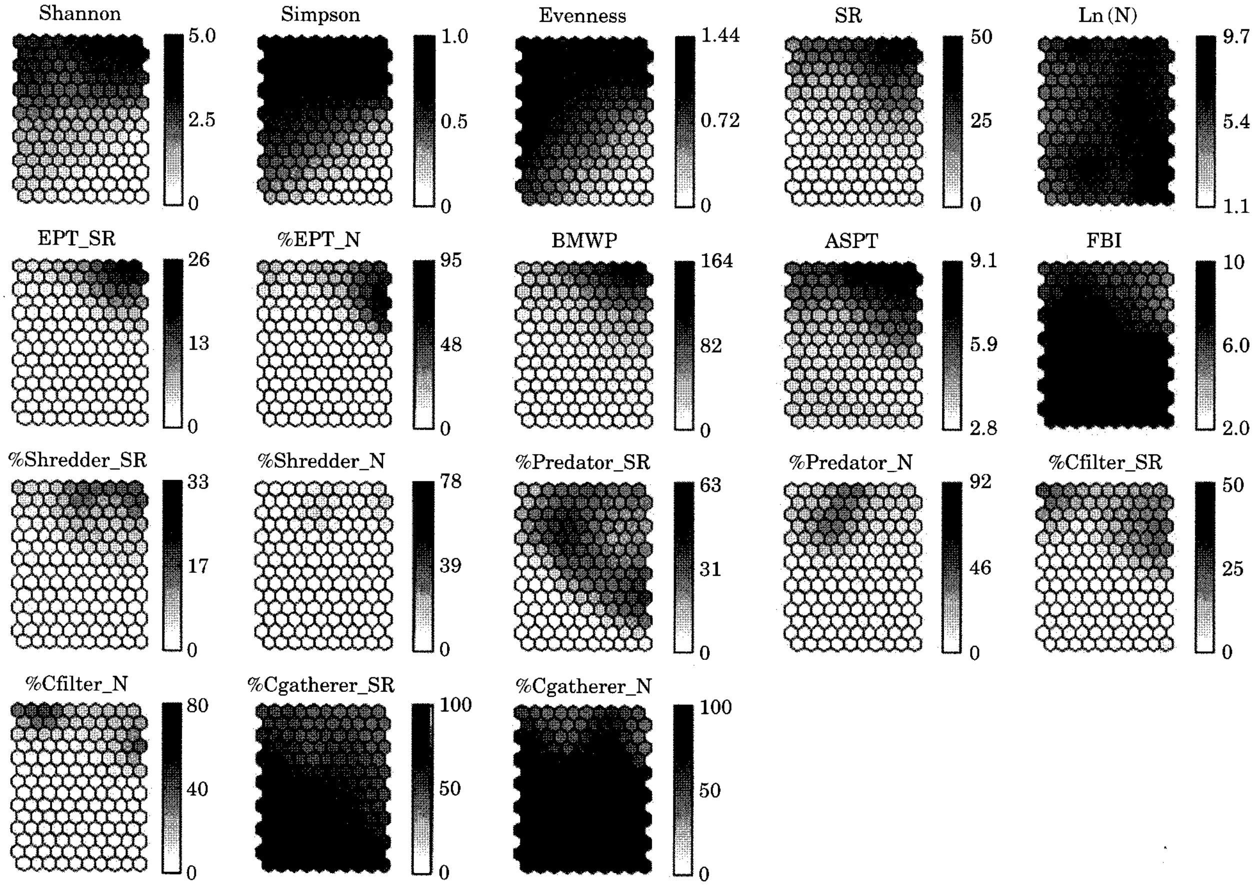

Fig. 4. Profile of biological indices corresponding to the clusters based on the trained SOM. The gray scale and the vertical bar indicate the gradients of each variable.

were even broader in their profiles with the gradient from the upper area to lower area of the SOM map (Fig. 4). In addition to high levels matching to cluster 4, both Shannon and Simpson diversity indices were higher at other clusters. The Simpson index was additionally higher at the left edge of the map, corresponding to cluster 5. The profile of evenness was similar with the Simpson index, showing high correlation ($r=0.95$, $p<0.001$).

While the gradients became weaker at the polluted area with the indices stated above, the indices based on saprobity showed stronger variation. Overall gradient clearly appeared with FBI (Fig. 4). The profile showed the vertical gradient, and the differences were distinctively compared with other indicators. Species richness (%) and abundance (%) in collector-gathers also showed the vertical gradients, showing high correlation with FBI (respectively $r=0.77$ and 0.87, $p<0.001$ for both).

Other indicators, however, did not show clear variation according to different pollution levels. Species richness (%) in predators did not display strong gradient of pollution levels, although it was relatively higher in clusters 3 and 5. It was notable that species richness (%) in predators was also partly high in the polluted areas. However, abundance (%) in predators was not high in cluster 5, and was only high in cluster 3 in the relatively clean area. Other indices related with FFGs were also uniquely presented on the map. Species richness (%) in collector-filterers generally matched to the gradient shown by biological indicators (e.g., total species richness ($r=0.87$, $p<0.001$) and ASPT ($r=0.88$, $p<0.001$)). Abundance (%) in collector-filterers was only specifically high at the top left corner of the map (Fig. 4), due to the exceptional occurrence of the dominant species, members of Hydropsychidae and Tanytarsini in cluster 1 (Table 3).

Environmental variables were accordingly dif-

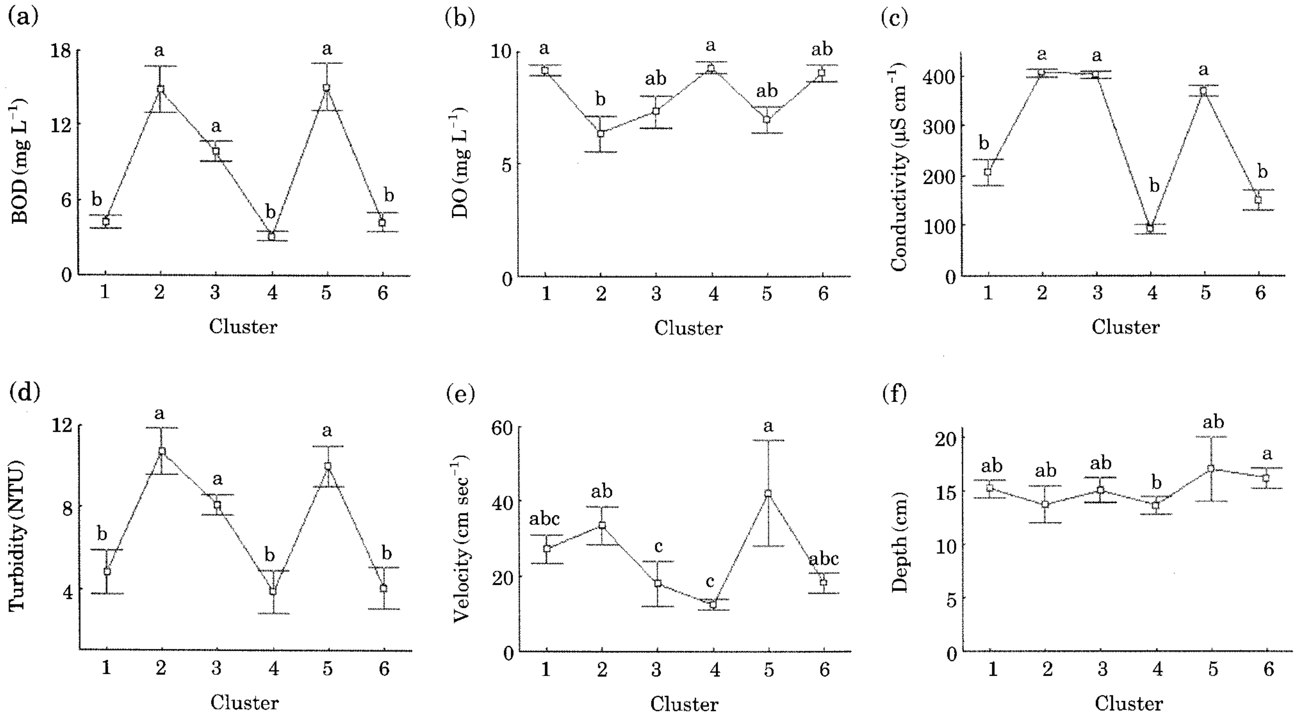

Fig. 5. The range of environmental variables in different clusters according to the SOM. (a) BOD, (b) DO, (c) conductivity, (d) turbidity, (e) velocity, (f) depth (significant differences based on Kruskal-Wallis test $p < 0.001$ except depth $p < 0.05$). Error bars indicate mean and standard error of each variable). Different alphabets indicate significant difference among clusters based on the multiple comparison test ($p < 0.05$).

Table 3. Community indices, key environmental factors and indicator species in different clusters

Cluster	Species richness	Abundance	Environments	Indicator species
1	Intermediate (15)*	Low	Low BOD, high DO	*Tanytarsus heusdensis* *Polypedilum cultellatum*
2	Low (5)	Intermediate	High BOD and turbidity	*Dicrotendipes pelochloris* *Limnodrilus hoffmeisteri*
3	Low (5)	High	Intermediate conductivity	*Physa acuta* *Tubifex tubifex*
4	High (29)	Intermediate	Low conductivity, pool zone	*Sweltsa nikkoensis* *Nigrobaetis bacilius*
5	Low (6)	Low	High velocity; deep	*Chironomus flaviplumus* *Erpobdella lineata*
6	Intermediate (19)	High	Low BOD and conductivity	*Acentrella sibirica* *Hydropsyche kozhantschikovi*

* average species richness per sample

ferent at different clusters defined in the SOM map (Fig. 5). BOD, conductivity and turbidity were correspondingly high at the clusters 2, 3 and 5 (K-W test, $p < 0.05$), while DO was in the lower range (mean 6.36~6.94 mg L^{-1}) at these clusters. This was in accordance with the variation in environmental variables listed on Table 1. In the clusters 2, 3, and 5, they were mainly consisted of the polluted sample sites from the Yangjae and Hakjang streams and two downstream sites (T3, T4) in the Soktae stream. Although species richness was all very low (mean 5.2 species per sample) in above 3 clusters, abundance was quite different among different clusters, in the range of 300~1,562 individuals per sample. In cluster 2, the average abundance was interme-

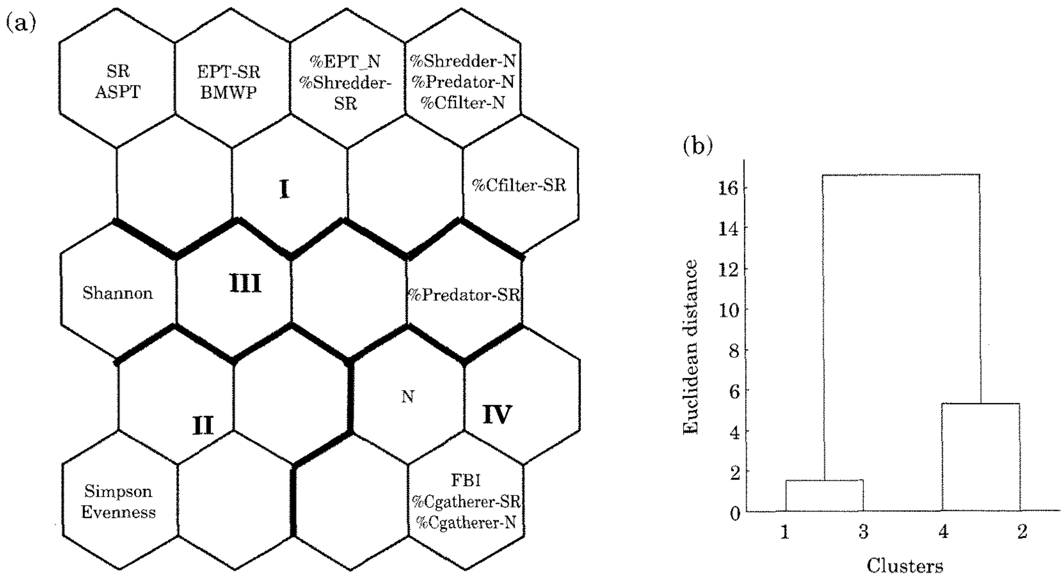

Fig. 6. (a) Association of biological indices according to the SOM. (b) Dendrogram of hierarchical cluster analysis with the Ward linkage method using Euclidean distance measure.

diate (average 750 individuals), and there were numerous collector-gatherers, while in the cluster 3 the species abundance were in the highest range in average (average 1,562 individuals). The abundance in cluster 5 was in the minimal range (average 300 individuals) compared with other clusters (Fig. 3). Also the indicator species in those three clusters were not the same (Table 3). The clusters 1, 4, and 6 generally represented clean area and showed the similar trends with their environmental condition (Fig. 5). However, velocity was relatively low in cluster 4, although they were not significantly different based on the multiple comparison test.

Samples in cluster 1 presented clean sites with low BOD and high DO. Abundance in communities was low while intermediate species richness was observed in this cluster (Table 3). This cluster was characterized with indicator species such as *Tanytarsus heusdensis* and *Polypedilum cultellatum*. Samples in cluster 2 were characterized with high values of BOD and turbidity and low value of species richness with indicator species *Dicrotendipes pelochloris* and *Limnodrilus hoffmeisteri*. Samples in cluster 3 were in the low species richness and high abundance in the intermediate conductivity condition, and were characterized with *Physa acuta* and *Tubifex tubifex*. Meanwhile, samples in cluster 4 showed high species richness at low values of conductivity. This cluster was characterized with indicator species including *Sweltsa nikkoensis* and *Nigrobaetis bacilius*. Cluster 5 represented the samples from highly polluted sites with high velocity, and showed low species richness and abundance. They were characterized with indicator species *Chironomus flaviplumus* and *Erpobdella lineata*. Finally, samples in cluster 6 were from low values of BOD and conductivity with intermediate species richness and high abundance, and characterized with indicator species such as *Acentrella sibirica* and *Hydropsyche kozhantschikovi* (Table 3).

Association among the biological indicators was also checked by using the SOM ordination in the R mode. The general trends shown in the profiles of the indicators in the Q mode analysis were reflected on association of the indicators (Fig. 6). The indices could be classified to 4 clusters based on the dendrogram of the hierarchical cluster analysis with the Ward linkage method. The indices in the highest range at the top right corner of the map corresponding to cluster 4 were grouped together on the map in cluster I. The frequently used indicators such as total species richness, EPT species richness, BMWP score, and ASPT were very closely located at the top right corner of the map in the same cluster ($r > 0.94$, $p < 0.001$). Majority of indices indicating proportion of FFGs were also grouped in the same cluster I. The Shannon index in cluster II was separated from the cluster representing the majority of biological indicators. The Simpson index and evenness index were also separated from the Shannon index belonging to cluster II (Fig. 6). The saprovity FBI and collector-gatherers were separately located in cluster IV on the map.

DISCUSSION AND CONCLUSIONS

The results in this study demonstrated the SOM was efficient in provision of comprehensive understanding on various biological indices proposed for water quality measurement in aquatic conditions. Biological indices commonly used for water quality, BMWP, ASPT, EPT, etc., were sensitively high in the clean area, specifically matching to cluster 4 (Figs. 3, 4). Most indices showed high correlations among indices as well as with their environmental condition, although indices behave differently in terms of identifying environmental disturbance according to the indices The association among the indices in this regard could be effectively presented with visualization of the estimated values on the SOM map. Generally, the values of indices showed strong negative correlations with pollution levels of the sample sites. Community indices such as Shannon and Simpson diversity indices, and total species richness showed the similar trends. In the comparison of biological indices, Friedrich *et al.* (1995) found high correlation between BMWP score and Saprobic index based on 232 samples from various stream types in Germany. Recently, Birk and Hering (2006) showed also similar results that assessment methods of the same type (Saprobic indices, BMWP/ ASPT scores) displayed the best correlation results. Gray and Delaney (2008) compared a range of benthic macroinvertebrate indices for the assessment of the impact of acid mine drainage on an Irish river, and found the differences in the ability of indices to discriminate the acid impacts with BMWP and Shannon indies which were the most precise. Taxa richness was also strongly correlated with acidic environmental parameters.

The results from our study confirmed the previous reports in the water quality indices. However, the present study showed additional information in the SOM analyses. The specific distributions of indices were differentiated in the SOM map regarding presentation of evenness in abundance. Consequently, diversity and evenness indices were different in presenting water quality from other biological indices such as BMWP, ASPT and EPT. The community indices tend to reflect degree of even occurrence of species more strongly. This was clearly presented by the profile of the Simpson and evenness indices: the values were strengthened in cluster 5 (Figs. 3, 4). In contrast, the Shannon index is more enhanced by species richness. The level of species richness was lower in the range around cluster 5, where Shannon index was more in accord with species richness (Fig. 4). This confirms the general characterization of two indices: The Shannon index is more strengthened by the number of species occurring at the sample sites, while the Simpson index is more influenced by the degree of evenness among the species (Magurran, 2004).

Differences among Shannon index, Simpson index and other biological indices were further confirmed by the R-mode analysis in the SOM and the Shannon index was separated from the cluster representing biological indicators (Fig. 6). Because the conventional diversity and evenness indices tended to be high even though the water quality state was poor representing relatively weak gradient at polluted sites, only using the conventional diversity and evenness may not represent the changes of water quality state. Most indices of FFGs showed also the similar patterns, excluding collector-gatherers. Shredders and collector-filterers were high in their species richness (%) in clean sample sites. While the conventional indices in general showed high values at the clean sites, the indices FBI, and species richness (%) and abundance (%) in collector-gatherers increased as the degree of pollution increased. It was notable that that the sample sites were mixed in the clusters, rather than reflecting geographic differences. Consequently, the indices more reflected the changes in water quality. It was also notable that the conventional biological indicators showed the shorter range in profiles: they were only high around the upper right corner of the SOM map and did not show the clear gradient. This implies that the index is oriented from the side of cleanness and is not much representative in the polluted side. The degradation of physical habitat and water quality can result in a reduction in the richness or number of intolerant macroinvertebrates (Benke *et al.*, 1981; Thorpe and Lloyd, 1999) that live in receiving streams (Roy *et al.*, 2003). The FBI, which is based on tolerance values of collected species, showed the high levels at the polluted area and presented the stronger gradient overall in the map. FBI appears to be suitable for indication of water quality: the index presented the degree of organic pollution by distinctly showing the vertical gradient (Fig. 4). This was confirmed with cluster 4 representing clean sites (Fig. 3). Collector-gatherers had high association with the FBI in the SOM map (Fig. 6), indicating the increase of proportion of collector-gatherers in the organic disturbed area. This is understandable that collector-gatherers occurred at the area with high amount of organic matters (Cummins and Wilzbach, 1985).

Classification of the samples based on the differences of the indices reflected the differences of environmental condition. Consequently, community compositions in different clusters were different. As indicated above, in different clusters, species richness showed gradient from polluted to clean sites, while showing the highest values in the intermediate disturbance condition. Abundance also displayed relation with species richness and degree of environmental disturbance. These results are supported by the intermediate disturbance hypothesis (Connell, 1978) assuming the high species diversity is a result of intermediate frequency of disturbance, while either too low or too high frequency of disturbance will result in a low biodiversity (Jørgensen and Padisak, 1996, Park *et al.*, 2003c).

In conclusion, the conveniently used biological indicators were efficiently patterned through the SOM. Conventional biological indices such as BMWP, ASPT and EPT appeared to be high at the clean areas, while the gradient was not strong at the polluted areas. Community diversity and evenness indices were different in presenting water quality from other biological indices such as BMWP, ASPT and EPT. The community indices tend to reflect degree of even occurrence of species more strongly. Our results also confirmed the general characterization of two indices: The Shannon index is more strengthened by the number of species occurring at the sample sites, while the Simpson index is more influenced by the degree of evenness among the species. Because the conventional diversity and evenness indices tended to be high even though the water quality state was poor representing relatively weak gradient at polluted sites, only using the conventional diversity and evenness may not represent the changes of water quality state. The gradient based on saprobity such as FBI showed the vertical gradient efficiently in illustrating the organic pollution.

ACKNOWLEDGMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2006-312-C00682). During the sample collecting and identification, we are indebt to all of laboratory members.

LITERATURE CITED

Alhoniemi, E., J. Himberg, J. Parhankangas and J. Vesanto. 2000. SOM Toolbox, http://www.cis.hut.fi/projects/somtoolbox.

Armitage, P.D., D. Moss, J.F. Wright and M.T. Furse. 1983. The performance of a new Biological Water Quality Score System based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Res.* **17**: 333-347.

Baptista, D.F., D. Buss, M. Egler, A. Giovanelli, M.P. Silveira and J.L. Nessimian. 2007. A multimetric index based on benthic macroinvertebrates for evaluation of Atlantic forest streams at Rio de Janeiro State, Brazil. *Hydrobiologia* **575**: 83-94.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA 841-B-99-002., U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Benke, A.C., G.G. Willke, F.K. Parrish and D.L. Stites. 1981. Effects of Urbanization on Stream Ecosystems. Report ERC07-81. Georgia Institute of Technology, Atlanta, GA, USA.

Birk, S. and D. Hering. 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water Framework Directive intercalibration exercise. *Hydrobiologia* **566**: 401-415.

Butcher, J.T., P.M. Stewart and T.P. Simon. 2003. A benthic community index for streams in the northern lakes and forests ecoregion. *Ecol. Indicat.* **3**: 181-193.

Chandler, J.R. 1970. A biological approach to water quality management. *Water Pollution Control* **69**: 415-422.

Chon, T.S., I.S. Kwak, M.Y. Song, Y.S. Park, H.D. Cho, M.J. Kim, E.Y. Cha, and S. Lek. 2002. Characterizing the effects of water quality on benthic stream macroinvertebrates in South Korea using a self-organizing mapping model, *In*: Ecology of Korea (Lee, D. ed.). Bumwoo Publishing Company, Seoul,

Korea.
Chon, T.S. and T.S. Kwon. 1991. Ecological studies on benthic macroinvertebrates in the Suyong River I. Investigations on community structures and biological indices in the lower reach. *Korea Journal of Limnology* **24**: 165-178.
Chon, T.S., Y.S. Park and J.H. Park. 2000. Determining temporal pattern of community dynamics by using unsupervised learning algorithms. *Ecological Modeling* **132**: 151-166.
Chon, T.S., Y.S. Park, K.H. Moon and E.Y. Cha. 1996. Patternizing communities by using an artificial neural network. *Ecol. Model.* **90**: 69-78.
Connell, J. 1978. Diversity in tropical rain forests and coral reefs. *Science* **199**: 1304-1310.
Cummins, K.W. and M.A. Wilzbach. 1985. Field Procedures for Analysis of Functional Feeding Groups of Stream Macroinvertebrates. Contribution 1611. Appalachian Environmental Laboratory, University of Maryland, Frostburg, Maryland.
Friedrich, G., E. Coring and B. Kuchenhoff. 1995. Vergleich verschiedener europä ischer Untersuchungs- und Bewertungsmethoden für Flieβgewasser. Landesumweltamt Nordrhein-Westfalen, Essen.
Gray, N.F. and E. Delaney. 2008. Comparison of benthic macroinvertebrate indices for the assessment of the impact of acid mine drainage on an Irish river below an abandoned Cu-S mine. *Environmental Pollution* **155**: 31-40.
Hawkes, H.A. 1979. Invertebrates as indicators of river water quality, *In*: Biological Indicators of Water Quality (James, A. and L. Evision eds.). John Wiley and Sons, Chichester, UK.
Hellawell, J.M. 1986. Biological Indicators of Freshwater Pollution and Environmental Management. Elsevier, London.
Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. *Great Lakes Entomology* **20**: 31-39.
Hilsenhoff, W.L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. *Journal of the North American Benthological Society* **7**(1): 65-68.
Hynes, H.B.N. 1960. The biology of polluted waters. Liverpool Univ. Press., London.
Jain, A.K. and R.C. Dubes. 1988. Algorithms for Clustering Analysis. Printice-Hall, Englewood Cliff, NJ.
James, A. and L. Evison. 1979. Biological Indicators of Water Quality. John Wiley & Sons, Ltd., NY.
Johnson, R.K. and W. Goedkoop. 2000. The 1995 national survey of Swedish lakes and streams: assessment of ecological status using macroinvertebrates, p. 229-240. *In*: Assessing the biological quality of freshwaters. RIVPACS and other techniques (Wright, J.F., D.W. Sutcliffe and M.T. Furse eds.). Freshwater Biological Association, Ambleside, UK.
Jørgensen, S.E. and J. Padisak. 1996. Does the intermediate disturbance hypothesis comply with thermodynamics? *Hydrobiologia* **323**: 9-21.
Kang, D.H., T.S. Chon and Y.S. Park. 1995. Monthly changes in benthic macroinvertebrate communities in different saprobities in the Suyong and Soktae Streams of the Suyong River. *Korean J. Ecol.* **18** (1): 157-177.
Kerrans, B.L. and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* **4**: 768-785.
KICT (Korea Institute of Construction Technology). 1997. Development of close-to nature river improvement techniques (CTNRIT) adapted to the Korean streams. Report, Vol. 2, KICT, Seoul.
KICT. 1999. Development of conservation, rehabilitation, and creation techniques of natural environment for the coexistence of man with nature. Development of close-to-nature river improvement techniques adapted to the Korean streams. Vol. I, KICT, Seoul.
Kohonen, T. 2001. Self-Organizing Maps. 3rd edition. Springer, Berlin.
Kohonen, T. 1989. Self-Organization and Associative Memory. Springer-Verlag, Berlin.
Kolkwitz, R. and M. Marsson. 1902. Grundsütze für die biologische Beurteilung des wassers nach seiner Flora und Fauda. Mitteilungen a. d. Kgl. Prüfungsanstalt f. *Wasserversorgung und Abwasserbeseitigung zu Berlin* **1**: 33-72.
Kolkwitz, R., M. Marsson. 1908. Ökologie der pflanzlichen Saprobien. *Ber. Dt. Bot. Ges.* **261**: 505-519.
Kwon, T.S. and T.S. Chon. 1991. Ecological studies on benthic macroinvertebrates in the Suyong River II. Investigations on distribution and abundance in its main stream and four tributaries. *Korea Journal Limnology* **24**: 179-198.
Magurran, A.E. 2004. Measuring biological diversity. Blackwell, Oxford.
Mielke, P.W., K.J. Berry and E.S. Johnson. 1976. Multi-response permutation procedures for a priori classifications. *Communications in Statistics-Theory and Methods* **5**: 1409-1424.
National Water Council (NWC), 1981. National Water Council, River Quality: The 1980 Survey and Future Outlook. National Water Council, London.
Oh, Y.N. and T.S. Chon. 1991a. A study on the benthic macroinvertebrates in the middle reaches of the Paenae stream, a tributary of the Nakdong River, Korea. I. Community analysis and biological assessment of the water quality. *Kor. J. Ecol.* **14**: 345-360.
Oh, Y.N. and T.S. Chon. 1991b. A study on the benthic macroinvertebrates in the middle reaches of Paenae stream, a tributary of the Nakdong River, Korea II. Comparison of communities and environments at the upper and lower sites of levees. *Kor. J. Ecol.* **14**: 399-413.
Oh, Y.N. and T.S. Chon. 1993. A study on the Benthic macroinvertebrates in the middle reaches of Pae-

nae stream, a tributary of the Nakdong River, Korea III. Drifting aquatic insects in four seasons. *Kor. J. Ecol.* **16**: 489-499.

Park, Y.S., R. Cereghino, A. Compin and S. Lek. 2003a. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecol. Model.* **160**: 265-280.

Park, Y.S., J. Chang, S. Lek, W. Cao and S. Brosse. 2003b. Conservation strategies for endemic fish species threatened by the Three Gorges Dam. *Conserv. Biol.* **17**: 1748-1758.

Park, Y.S., T.S. Chon, I.S. Kwak, J.K. Kim and S.E. Jørgensen. 2001. Implementation of artificial neural networks in patterning and prediction of exergy in response to temporal dynamics of benthic macroinvertebrate communities in streams. *Ecol. Model.* **146**: 143-157.

Park, Y.S., T.S. Chon, I.S. Kwak and S. Lek. 2004. Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. *Sci. Total Environ.* **327**: 105-122.

Park, Y.S., M.Y. Song, Y.S. Park, K.H. Oh, E.C. Cho and T.S. Chon. 2007. Community patterns of benthic macroinvertebrates collected on the national scale in Korea. *Ecol. Model.* **203**: 26-33.

Park, Y.S., P.F.M. Verdonschot, T.S. Chon and S. Lek. 2003c. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. *Water Research* **37**(8): 1749-1758.

Pielou, E.C. 1966. Shannon's formulae as a measure of specific diversity: its use and misuse. *American Naturalist* **100**: 463-465.

Qu, X.D., M.Y. Song, Y.S. Park, Y.N. Oh and T.S. Chon. 2008. Species abundance patterns of benthic macroinvertebrate communities in polluted streams. *Annales de Limnologie-International Journal of Limnology* **44**: 119-133.

Rosenberg, D.M. and V.H. Resh. 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman & Hall, London.

Roy, A.H., A.D. Rosemond, M.J. Paul, D.S. Leigh and J.B. Wallace. 2003. Stream macroinvertebrate response to catchment urbanisation (Georgia, U.S.A.). *Freshwater Biology* **48**: 329-346.

Siegel, S. and N.J. Castellan. 1988. Nonparametric Statistics for the Behavioural Sciences., McGraw-Hill., New York.

Sladecek, V. 1969. The measures of saprobity. *Verh. Int. Ver. Limnol.* **16**: 809-816.

Sladecek, V. 1979. Continental systems for the assessment of river water quality. *In*: Biological Indicators of Water Quality (James, A. and L. Evison eds.). John Wiley & Sons, Chichester.

Song, M.Y., H.J. Hwang, I.S. Kwak, C.J. Ji, Y.N. Oh, B.J. Youn and T.S. Chon. 2007. Self-organizing mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation. *Ecological Modelling* **203**: 18-25.

Song, M.Y., S.E. Lee, J. Park, J. Park, B. Kim, S. Koh, K. Lee, Y.S. Park and T.S. Chon. 2005. Comparative community analysis of benthic macroinvertebrates and microorganisms across different levels of organic pollution in a stream by using artificial neural networks. *WSEAS Transactions on Biology and Biomedicine* **3**(2): 257-268.

Song, M.Y., Y.S. Park, I.S. Kwak, H.S. Woo and T.S. Chon. 2006. Characterization of benthic macroinvertebrate communities in a restored stream by using self-organizing map. *Ecol. Infor.* **1**: 295-305.

Spellerberg, I.F. 1991. Monitoring Ecological Change. Cambridge University Press, Cambridge.

StatSoft, Inc., 2004. STATISTICA (data analysis software system), version 7, www.statsoft.com.

The Mathworks, 2001. The Mathworks Inc., 2001. MATLAB Version 6.1, Massachusetts.

Thorpe, T. and B. Lloyd. 1999. The macroinvertebrate fauna of St. Lucia elucidated by canonical correspondence analysis. *Hydrobiologia* **400**: 195-203.

Vesanto, J. and E. Alhoniemi. 2000. Clustering of the self-organizing 12 map. *IEEE Transactions on Neural Networks* **11**: 586-600.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**: 236-244.

Wilhm, J. 1968. Biomass unites versus numbers of individuals in species indices. *Ecology* **49**: 153-156.

Wilhm, J. 1972. Graphic and mathematical analyses of biotic communities in polluted streams. *Ann. Rev. Entomol.* **17**: 223-252.

Winget, R.N. and F.A. Mangun. 1979. Biotic Condition Index: Integrated Biological, Physical and Chemical Stream Parameters for Management. US Forest Service, Intermountain region, Provo, Utah.

Woodwiss, F.S. 1978. Report of U.K. Participants in the E.E.C.

Youn, B.J. and T.S. Chon. 1996. Community analysis in chironomids and biological assessment of water qualities in the Suyong and Soktae streams of the Suyong River. *Korea Journal Limnology* **29**: 275-289.

(Manuscript received 1 August 2009,
Revision accepted 13 September 2009)