

Ecological informatics approach to screening of integrity metrics based on benthic macroinvertebrates in streams

Woon-Seok Cho¹, Young-Seuk Park², Hae-Kyung Park³, Hak-Yang Kong³ and Tae-Soo Chon^{1*}

¹ Department of Biological Sciences, Pusan National University, Busan 609-735, Republic of Korea

² Department of Biology, Kyung Hee University, Seoul 130-701, Republic of Korea

³ Water Environment Research Department, The National Institute of Environmental Research, Incheon 404-170, Republic of Korea

Received 30 August 2010; Accepted 6 March 2011

Abstract – Benthic macroinvertebrates are considered as a representative taxon that indicates the ecological status of freshwater ecosystems. Numerous indices derived from community data have been proposed to estimate either biological water quality or ecosystem health. In this study, metrics based on benthic macroinvertebrates at the family level were screened using ecological informatics to provide a multi-metric measurement that would be suitable for presenting ecological integrity across different levels of environmental impact. Benthic macroinvertebrates were collected at a total of 720 sample sites from river basins and streams in Korea in 2009. Based on four categories of community status (*i.e.*, diversity, richness, tolerance, and composition), 37 metrics were selected as initial candidates according to the literature. The candidate metrics were evaluated according to parameters including discriminatory power, redundancy, and responsiveness to stressors. Self-organizing map was utilized to assist the screening procedure by providing ordination, clustering, and visualization of metric and environmental data. Six metrics were finally selected as a multi-metric and were compared with conventional indicators for presenting the ecological integrity of streams.

Key words: Multi-metric / ecological integrity / biological indicator / aquatic ecosystem / discrimination power

Introduction

Benthic macroinvertebrate communities play a key role in the functioning of aquatic ecosystems by filling the gap between producers (*i.e.*, benthic algae) and higher consumers (*i.e.*, fishes). Characteristics including high diversity, sedentariness of life style, and a relatively long life span (*i.e.*, season–year) in benthic macroinvertebrates make them suitable for continuous monitoring of ecological status in an integrative manner. Benthic macroinvertebrates have been used as efficient indicator taxa for ecological evaluation of environmental impact in aquatic ecosystems (Hawkes, 1979; Rosenberg and Resh, 1993; Barbour *et al.*, 1996; Reynoldson *et al.*, 1997; Wright *et al.*, 2000). Various biological indices have been developed to present biological water quality including the Hilsenhoff biotic index (Hilsenhoff, 1987), invertebrate community index (Karr & Chu, 1999), biological monitoring working party (BMWP) (Walley and Hawkes, 1996, 1997), average score per taxon (Armitage *et al.*, 1983), and Ephemeroptera, Plecoptera, and Trichoptera (EPT)

richness (Lenat, 1988). Due to the high diversity and abundance of benthic macroinvertebrates, numerous metrics have been proposed to estimate ecological integrity (Lenat, 1993; Barbour *et al.*, 1995, 1996; Reynoldson *et al.*, 1997; Blocksom *et al.*, 2002; Klemm *et al.*, 2002). Up to 237 metrics (Whittier *et al.*, 2007) have been reported and accordingly evaluated in streams (Kerans and Karr, 1994; Barbour *et al.*, 1999; Klemm *et al.*, 2002; Blocksom, 2003; Böhmer *et al.*, 2004; Ode *et al.*, 2005; Whittier *et al.*, 2007; Stoddard *et al.*, 2009) and in lakes (Blocksom *et al.*, 2002; Solimini *et al.*, 2008; Taowu *et al.*, 2008; Trigal *et al.*, 2009). The criteria used for selection were the degree of gradient (*i.e.*, presenting the overall impact of an environment), discriminatory power (Blocksom *et al.*, 2002; Klemm *et al.*, 2003), and the decrease in redundancy among similar metrics (Klemm *et al.*, 2003).

One difficulty in using metrics, however, lies in their multitude: a large number of metrics should be decreased to produce a small, but practical number of metrics to present ecological integrity more promptly and effectively. In this study, we demonstrated a screening procedure using ecological informatics. The self-organizing map (SOM) was utilized for extracting information from the

*Corresponding author: tschon@pusan.ac.kr

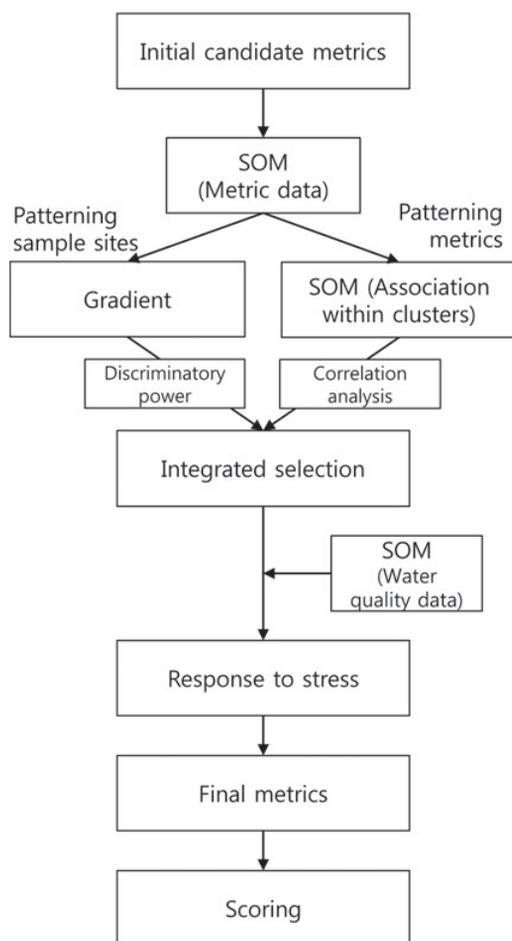


Fig. 1. Screening procedure of a multi-metric based on benthic macroinvertebrates in streams.

metric data. The SOM has been an efficient tool for clustering and visualization of complex data without prior knowledge, and has been extensively used for patterning communities of benthic macroinvertebrates since the 1990s (Chon *et al.*, 1996; Park *et al.*, 2003; Song *et al.*, 2007). The SOM was utilized for assisting with the screening procedure, covering discriminatory power, estimating the degree of association, and evaluating responsiveness to environmental impact. In this study, a multi-metric is proposed to present ecological integrity in streams across different levels of pollution.

Materials and methods

Process of metric evaluation

Metrics screening was conducted under guidelines to secure objectivity in presenting ecological states, and to achieve simplicity and promptness in evaluation. The overall selection procedure is presented in a flow chart (Fig. 1), which displays the discriminatory power estimation (Barbour *et al.*, 1996), for determining degree of association and checking responsiveness to stressors (Blocksom *et al.*, 2002). To extract information from

community and environmental data, the SOMs were applied three times during the screening procedure for: (1) ordination of metrics for assisting the gradient evaluation, (2) checking redundancy (*i.e.*, degree of association) among metrics through clustering, and (3) evaluation of responsiveness to water quality data based on visualization (Fig. 1). Detailed procedures will be discussed along with the step-by-step screening processes presented in the Results section.

Community data

The data used for measuring metrics were obtained from a research project “the Survey and Evaluation of Aquatic Ecosystem Health in Korea” supported by The Ministry of Environment and National Institute of Environmental Research, Korea. A total of 720 sites were selected in streams including the four largest river watersheds in Korea: Han River (320 sites), Nakdong River (130 sites), Keum River (130 sites), and Yeongsan–Seomjin Rivers (140 sites). Benthic macroinvertebrates were surveyed two times separately in spring (April) and autumn (September) in 2009. Samples were collected with a Surber sampler (30 × 30 cm²; mesh size 1 mm), in three replications, in riffle areas of the streams. After fixation in 95% ethyl alcohol in the field, the samples were transported to the laboratory for sorting and identification, and were stored in 80% ethyl alcohol. Densities were converted to unit area (1 m²). Community indices including species richness, diversity (Shannon and Weaver, 1949), evenness (Pielou, 1975), dominance (McNaughton, 1967), and biological indicators such as Korean saprobic index (KSI) (Won *et al.*, 2006) and the revised BMWP (Walley and Hawks, 1997) were measured based on the community data.

Water quality indicators including biochemical oxygen demand (BOD), NH₃-N, NO₃-N, total nitrogen (TN), and PO₄-P, total phosphorus (TP) were also obtained from the same sample sites according to the guidelines of the “National Surveys for Stream Ecosystem Health” in Korea (MOE/NIER, 2008).

Initial candidate metrics

Considering that classification of macroinvertebrates to the levels of genus and species requires an exceptionally high level of effort (*i.e.*, delay in classification), we focused on family-based metrics in this study, and emphasized promptness and objectivity in evaluation of ecological integrity *in situ*. Based on the literature (Barbour *et al.*, 1999; Blocksom *et al.*, 2002; Klemm *et al.*, 2002; Klemm *et al.*, 2003; Böhmer *et al.*, 2004; Ode *et al.*, 2005; Hargett and Zumberge, 2006; Purcell *et al.*, 2009), 37 metrics were initially chosen from community data of benthic macroinvertebrates (Table 1). Kerans and Karr (1994) and Klemm *et al.* (2003) proposed five categories of metrics including richness, tolerance, composition, functional

Table 1. Initial candidate metrics of benthic macroinvertebrates based on the literature for selection of a multi-metric.

Class	Abbreviation	Metric	Response to stress*	Literature**
Diversity	SMDIV	Simpson diversity	–	2, 3, 8
	SHDIV	Shannon diversity	–	2, 3, 4, 5, 8
	EVN	Evenness	–	8
	DOM1	Dominance of 1st family	+	1, 2, 5, 6, 7, 8
Richness	DOM2	Dominance of 1st and 2nd families	+	2, 8
	FAM	Number of families	–	1, 2, 5, 6, 7, 8
	FEPT	Number of EPT families	–	1, 2, 3, 4, 5, 6, 7, 8
	FEPH	Number of Ephemeroptera families	–	1, 2, 3, 5, 6, 7, 8, 8
	FPLE	Number of Plecoptera families	–	1, 2, 3, 4, 5, 6, 7, 8
	FTRI	Number of Trichoptera families	–	1, 2, 3, 4, 5, 6, 7, 8
	FCRU	Number of Crustacea families	–	3
	FMOL	Number of Mollusca families	–	3
	% FEPT	Percent of EPT families	–	2, 6, 7, 8
	% FEPH	Percent of Ephemeroptera families	–	2, 6, 7, 8
	% FPLE	Percent of Plecoptera families	–	2, 6, 7, 8
	% FTRI	Percent of Trichoptera families	–	2, 3, 6, 7, 8
	% FCRU	Percent of Crustacea families	–	3
	% FMOL	Percent of Mollusca families	–	3
Tolerance	FBI	Family biotic index	+	1, 2, 6, 8
	FINT	Number of intolerance families	–	1, 2, 3, 5, 6, 8
	FTOL	Number of tolerance families	+	5, 8
	FFAC	Number of facultative families	±	8
	% FINT	Percent of intolerance families	–	2, 6, 8
	% FTOL	Percent of tolerance families	+	2, 3, 5, 6, 8
	% FFAC	Percent of facultative families	±	2, 8
	% NINT	Percent of intolerance individuals	–	5, 8
	% NTOL	Percent of tolerance individuals	+	1, 3, 5, 8
	% NFAC	Percent of facultative individuals	±	8
Composition	% NEPT	Percent of EPT individuals	–	1, 3, 4, 5, 6, 8
	% NEPH	Percent of Ephemeroptera individuals	–	1, 2, 3, 5, 6, 7, 8
	% NPLE	Percent of Plecoptera individuals	–	2, 3, 4, 5, 6, 7, 8
	% NTRI	Percent of Trichoptera individuals	–	2, 5, 6, 7, 8
	% NCHI	Percent of Chironomidae individuals	+	1, 2, 5, 6, 7, 8
	% NNINS	Percent of non-insect individuals	±	3, 6, 8
	% NCRU	Percent of Crustacea individuals	±	3, 4, 5
	% NOLI	Percent of Oligochaeta individuals	+	5, 6, 7, 8
	% NMOL	Percent of Mollusca individuals	±	5

* +, positive; –, negative; ±, both.

**1: Barbour *et al.* (1999), 2: Klemm *et al.* (2002), 3: Klemm *et al.* (2003), 4: Böhmer *et al.* (2004), 5: Ode *et al.* (2005), 6: Hargett and ZumBerge (2006), 7: Purcell *et al.* (2009), and 8: Blocksom *et al.* (2002).

feeding group, and habitat type, while Stoddard *et al.* (2009) classified candidate metrics in six categories with the addition of diversity. In this study, we chose four categories: diversity, richness, tolerance, and composition, because functional feeding groups and habitat types frequently overlapped within the family level (*e.g.*, Baetidae presenting scraper and collector–gatherer at the same time). In each class, degree of responsiveness to disturbance, simplicity of measurement, and data availability were further considered as additional criteria for choosing the candidates. Additionally, direction of response to stress was also considered; indicators showing both positive (*e.g.*, family biotic index (FBI), percent of Chironomidae individuals), and negative (*e.g.*, BMWP, Shannon diversity index) responses to stressors were included together in the list (Table 1).

SOM

We utilized the SOM for patterning of macroinvertebrate communities (Ward, 1963; Park *et al.*, 2003). The output layer consists of $L \times M$ computation nodes in the SOM. Each neuron is represented as j ; the output layer is arranged in two dimensions for convenience of visual understanding. Suppose community data containing S species and the density of species (i) is expressed as a vector (\mathbf{x}_i). The vector (\mathbf{x}_i) is considered to be an input layer to the SOM. In the network, each node (j) of the output layer is connected to each node (i) of the input layer. The connectivity is represented as the weights, $w_{ij}(t)$, which adaptively change in each iteration of calculation, (t). Initially, the weight is randomly assigned in small values. Each neuron of the network computes

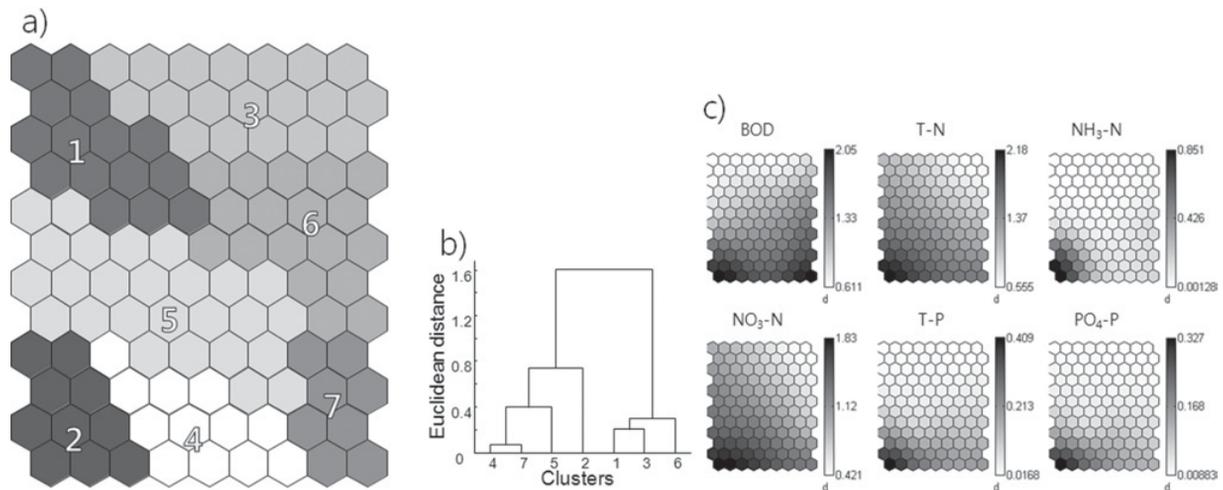


Fig. 2. Clustering of the sampling sites on the SOM based on water quality indicators. (a) Clusters, (b) dendrogram based on Ward's linkage method, and (c) profiles of water quality indicators.

the summed distance between the weights, and the distance $d_j(t)$ at the output node (j) is calculated as shown below:

$$d_j(t) = \sum_{i=0}^{S-1} (x_i - w_{ij}(t))^2. \quad (1)$$

The neuron responding maximally to a given input vector was chosen to be the winning neuron, the weight vector of which has the shortest distance to the input vector. The winning neuron, and possibly its neighboring neuron(s), is allowed to learn by changing the connecting weights in a manner to further reduce the distance between the weight and the input vector as shown below:

$$w_{ij}(t+1) = w_{ij}(t) + a(t)(x_i - w_{ij}(t)Z_j), \quad (2)$$

where Z_j is assigned to 1 for the winning (and its neighboring) neuron(s), while it is assigned 0 for the rest neurons, and $a(t)$ denotes the fractional increment of the correction.

After training, the Ward's linkage method (Ward, 1963) was applied to the weights of the SOM for clustering the patterned nodes. The initialization and training processes followed suggestions by the SOM Toolbox by allowing optimization in a logarithm (Zurada, 1992; Chon *et al.*, 1996; Vesanto *et al.*, 2000) developed by the Laboratory of Information and Computer Science at the Helsinki University of Technology (<http://www.cis.hut.fi/projects/somtoolbox/>) under Matlab environments (The Mathworks Inc., 2001). A detailed description regarding application of the SOM to ecological data is provided by Park *et al.* (2003).

The input values having greatly different values in densities are avoided for training. The data were transformed by a natural logarithm to emphasize the differences in low densities. Subsequently, the transformed data were proportionally normalized between 0 and 1 in the range of the maximum and minimum density for each species collected during the survey period.

Results

Environmental factors

Differences of environmental factors (mainly consisting of water quality indicators) measured in the sampling sites were patterned by the SOM (108 = 12 × 9 nodes) (Fig. 2a). The clusters were formed according to different levels of pollution along with the vertical gradient. Based on the Ward's linkage method (Ward, 1963) seven clusters were observed (Fig. 2b). The less polluted sites were grouped in clusters 1 and 3 in the upper area, while the polluted sites were placed in the bottom area in clusters 2, 4, 5, and 7 (Fig. 2a).

The profiles of water quality indicators were visualized on the SOM (Fig. 2c); BOD, TN and NO₃-N spanned a broad range in the map. BOD tended to show the vertical gradient, while the water quality factors such as TN, and NO₃-N presented the diagonal gradients. The values for NH₃-N and PO₄-P, however, appeared only in a limited range on the map (Fig. 2c).

Metrics clustering

The initial 37 candidate metrics were used to form an input data matrix to be trained with the SOM in two ways: patterning sample sites (variables; 37 metrics, and cases; sample sites), and revealing association of metrics by transposing the data matrix (Fig. 1).

Gradient

Similar to the case of environmental factors (Fig. 2a), the sampling sites were classified based on the metrics (Figs. 3a and 3b), and the classification revealed differences of their pollution levels. A division was observed between the clusters presenting the severely polluted sites (clusters 2, 6, and 7) in the bottom area, and the clusters presenting the intermediately polluted sites (clusters 4

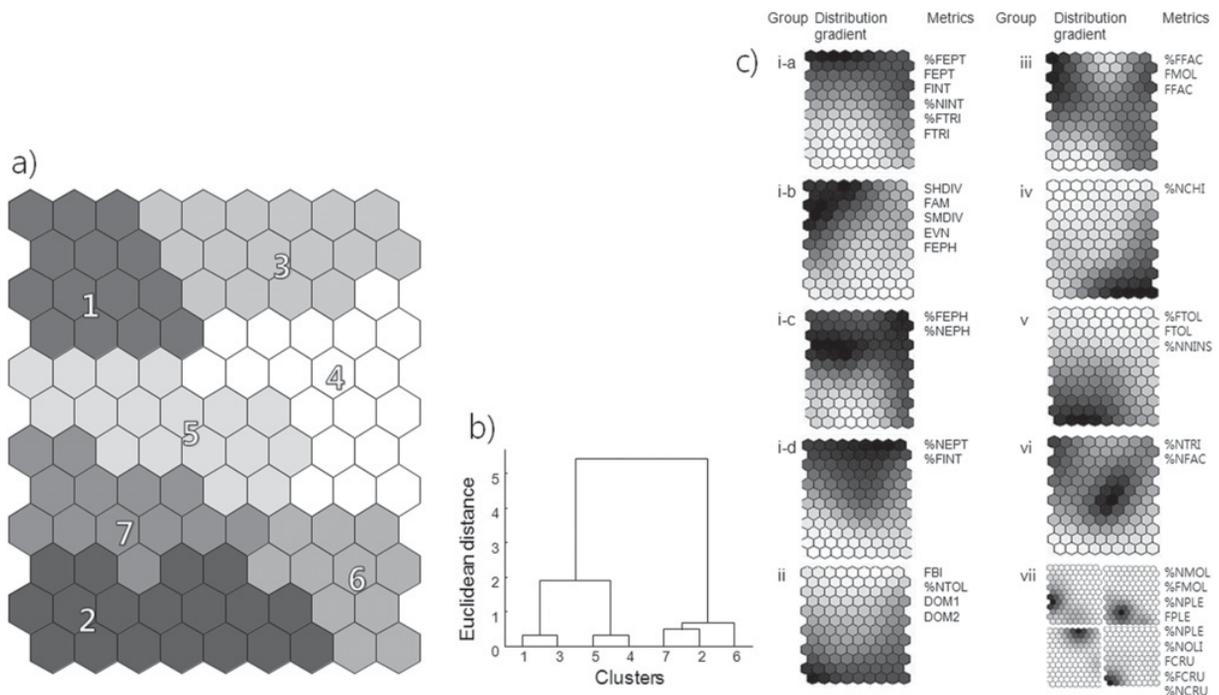


Fig. 3. Clustering of the sampling sites on the SOM based on the metrics. (a) Clusters, (b) dendrogram based on Ward’s linkage method, and (c) profiles of metrics in different groups (SOMs for group vii in Fig. 3c are examples for metrics belonging to this group).

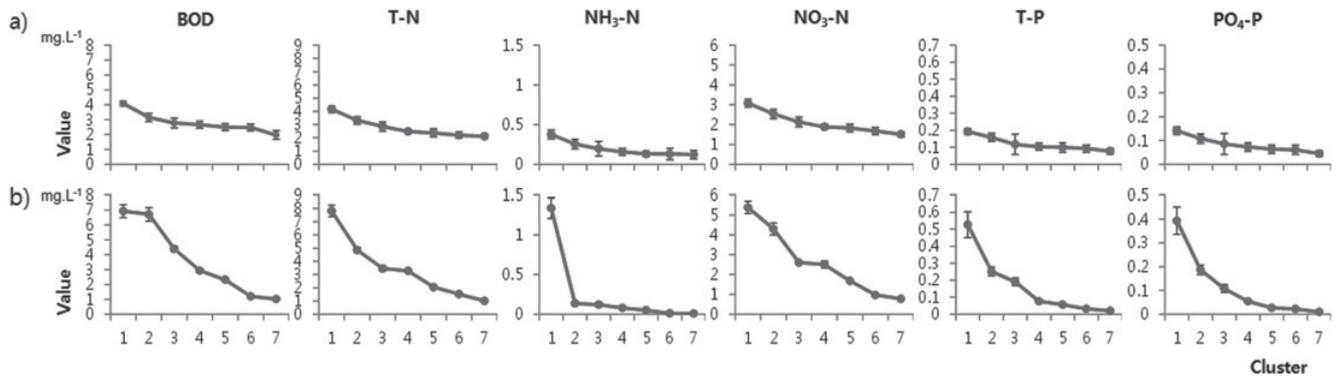


Fig. 4. Differences in water quality indicators across different clusters on the SOM trained with metrics (a) and water quality indicators (b). Error bar indicates standard error.

and 5) in the middle area of the map. Each cluster was presented by different groups of metrics. Cluster 1, for instance, accommodated the largest group of the sampling sites with high diversity and species richness, whereas cluster 6 covered communities with comparatively low diversity indices. The clusters were also characterized by different taxa. Plecoptera, Ephemeroptera, and Trichoptera were abundant in clusters 1–5, whereas Oligochaeta and Chironomidae were more collected in clusters 2 and 6. Mollusca were frequently observed in cluster 7.

For comparison of SOM classification in metric and water quality factors, the number of clusters was fixed to seven clusters in each dataset. The order of clusters was arranged according to the order of high value (*e.g.*, level of water quality indicators in the *y*-axis in Fig. 4). The gradients in metrics and water quality indicators were

accordingly observed (Fig. 4). As expected, the gradients of water quality indicators were observed when the same variables were visualized across the clusters (Fig. 4b). It was notable that the gradient in metric (Fig. 4a) was also similar to the gradient shown in the water quality indicators. This indicated that the SOM patterning based on metrics and water quality factors showed similar trends in presenting the level of environmental impact.

Discriminatory power

Profiles of the metric data were further visualized on the SOM using weight vectors of each metric in the trained SOM (Fig. 3c). Dark gray indicates high value for each metric, whereas light one presents low value. Values were rescaled in minimum and maximum to

Table 2. Overall procedure and list of the screened metrics in evaluation of discrimination power, degree of association, and response to stress.

Class	Metrics					Response to environmental factors	
	Association	Discrimination power	Score	Combined evaluation	Increase/decrease	Sensitivity	Multi-metric
Diversity	SMDIV		2	SMDIV	–	sensitive	SMDIV
	SHDIV		2				
	EVN		1				
	DOM1		2				
	DOM2		2				
Richness	FAM		2				
	FEPT	FEPT	3	FEPT	–	most sensitive among richness class	FEPT
		FEPH	3				
	FPLE		3				
	FTRI	FTRI	3	FTRI	–	less sensitive than FEPT	
	% FEPT	% FEPT	3	% FEPT	–	less sensitive than FEPT	
	% FPLE		3				
Tolerance	FBI		2				
	FINT	FINT	3	FINT	–	less sensitive than % FINT	
	FTOL	FTOL	3	FTOL	+	less sensitive than % FTOL	
		FFAC	0				
	% FINT	% FINT	3	% FINT	–	most sensitive to stress	% FINT
	% FTOL	% FTOL	3	% FTOL	+	less sensitive than % FTOL	
	% NINT		0				
	% NTOL	% NTOL	3	% NTOL	+	better than % FTOL	% NTOL
Composition	% NEPT	% NEPT	3	% NEPT	–	less sensitive than FEPT	
	% NEPH	% NEPH	3	% NEPH	–	most sensitive in composition class	% NEPH
		% NTRI	2				
		% NCHI	2				

visualize distribution gradient. We made seven groups of candidate metrics based on their distribution gradient patterns for the selection of metric for further screening. Vertical distribution (groups i and ii) was initially considered for showing a higher level of discriminatory power explained in the next section. Horizontal (group iii) and diagonal (groups iv and v) distributions could be secondarily considered for selecting the candidates. Some metrics such as % NMOL and % FMOL were only in a limited area (group vii), whereas two metrics % NTRI and % NFAC (group vi) distributed widely without showing clear distribution gradient.

Discriminatory power is defined as the ability to discriminate between unpolluted and polluted sites (*i.e.*, showing wider and higher separation). The scoring was proposed by Barbour *et al.* (1996). The degree of impact was scored 0 to 3, from minimum to maximum, in each metric according to a Box-and-Whisker plot (Barbour *et al.*, 1996; Blocksom *et al.*, 2002). Scoring of discriminatory power has been applied to macroinvertebrates (Blocksom *et al.*, 2002; Klemm *et al.*, 2002). When the metrics were measured with the sampled data in between maximum and minimum values, interquartile ranges could be determined. If the metric showed no overlap of interquartile ranges between unpolluted and polluted sites, “3” was scored as the maximum discriminatory power. When both medians of polluted or unpolluted sites were outside of interquartile ranges, even though interquartile

ranges were overlapped, “2” was scored. If only one median of polluted or unpolluted sites was outside of interquartile ranges, “1” was scored. When both medians of polluted or unpolluted sites were inside of interquartile ranges, “0” was scored (Blocksom *et al.*, 2002).

In this study, discriminative power was applied to values of metrics visualized on the SOM (Fig. 3c). According to the scoring system, metrics that scored “3” in discriminatory power, or most highly scored metrics in the same class, were selected for further evaluation. We used the KSI (Won *et al.*, 2006) as the standard indicator for the reference index. Unpolluted sites corresponding to an oligosaprobic condition were <0.7 in KSI values, while polluted sites were >3.1, being equivalent to α -mesosaprobic conditions in a saprobity range while the maximum value was up to 4.6 (Won *et al.*, 2006). According to the observed values in determining water quality in the minimal and maximal ranges, the sampling sites with maximal KSI values ranging from the 75 percentile to the 100 percentile were defined as polluted sites, while the sampling sites presenting the KSI values less than the 25 percentile were considered as unpolluted sites.

Within each group of metrics shown in Figure 3c, the metrics were arranged in the order of high scores based on discriminatory power. The class of metrics presenting diversity, richness, tolerance, and composition were considered together for screening. The list of the selected metrics is presented in Table 2.

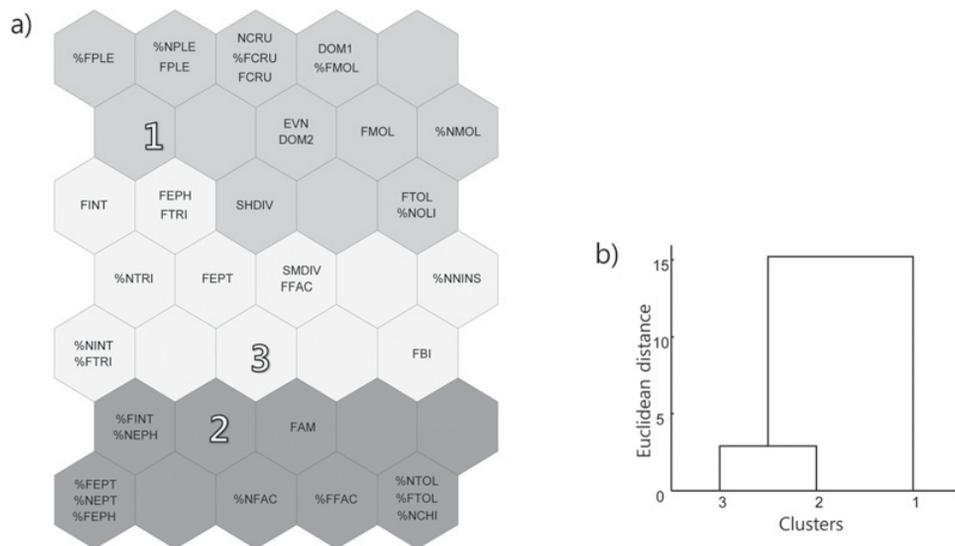


Fig. 5. Clustering of metrics on the SOM trained with the transposed dataset used in Figure 3. (a) Clusters and (b) dendrogram based on Ward's linkage method.

Association of metrics

In addition to patterning of the sampling sites, the metrics were classified by the SOM to reveal degree of association among metrics (Fig. 5). Considering that the metrics presented within each cluster would be similar, either one or several metrics could be selected to represent the whole group of metrics within each cluster. Using this method, redundancy among the similar metrics could be minimized. Correlation coefficients between the metrics were used to screen metrics within the clusters (Table 3). The metrics showing the highest ranges of correlations with other metrics were selected as the candidates. Relatively higher levels of correlation coefficients were observed in clusters 2 and 3 (Table 3). The metrics presenting the majority of correlation coefficients >0.4 were chosen in this case. In cluster 1, the metrics, however, showing correlation coefficients >0.2 were selected as candidates for the next procedure, because correlation coefficients were generally lower in this cluster compared with other clusters. Through this process, 20 metrics were chosen, including number of families (FAM), FBI, Shannon diversity (SHDIV), and percent of tolerance families (% FTOL) (Table 2).

Integrated evaluation

After the selection of metrics through SOM patterning, the metrics were subsequently screened after combining two datasets (*i.e.*, gradient and association) (Fig. 1). The metrics selected through this process mostly showed the highest score of “3” as their discriminating power (Table 2). In addition to these metrics, one index, Simpson diversity (SMDIV), was chosen to represent the class of diversity, because metrics in diversity were not selected for discriminating power (Table 2). Direction of response to stress (*i.e.*, positive or negative) was also presented together in the list.

Response to stress

The selected metrics were additionally screened according to degree of responsiveness to environmental impact (*i.e.*, compatibility to physico-chemical factors) (Fig. 1). For this purpose, water quality indicators patterned by the SOM were used as a reference for selecting metrics. After training with water quality indicators, the metric data were superposed according to different clusters based on the SOM (Fig. 6). Statistical differences of the metric values in different clusters were evaluated through a multiple comparison with the Tukey HSD test. Metrics showing the higher degree of significance between clusters were regarded as more responsive to environmental impact (Fig. 6). For example, the number of EPT families (FEPT) was significantly different in four groups (A–D), while the percent of EPT families (% FEPT) was different in three groups (A–C). Consequently, FEPT was regarded as being more responsive to environmental impact than % FEPT. Through this process, five metrics were finally selected to form a new multi-metric as listed in Table 2: percent of intolerance families (% FINT), percent of tolerance individuals (% NTOL), FEPT, SMDIV, and percent of Ephemeroptera individuals (% NEPH). A new scoring system was coined as macroinvertebrate integrity score in stream (MISS), based on the screening procedure presented in this study.

Scoring and evaluation

After final screening, actual MISS scores could be given from field data. There are two criteria in scoring multi-metrics (Simon and Lyons, 1995; Blocksom *et al.*, 2002): (a) setting value according to reference conditions or based on distribution across all sampling sites (Simon and Lyons, 1995); and (b) converting the value of metrics either by discrete values of “1, 3, or 5” (Kerans and Karr, 1994) or by continuous values of “0–1” (Blocksom, 2003)

Table 3. Correlation coefficients between metrics within clusters defined in the SOM in Figure 5. Metrics in bold are the selected metrics for the next step of evaluation.

Clusters	Variables	% NCRU	% FMOL	% NMOL	% NOLI	% FPLE	% NPLE	FCRU	DOM2	DOM1	EVN	FMOL	FPLE	SHDIV	FTOL	
Cluster 1	% FCRU	0.71**	0.12**	0.04	0.00	-0.07	-0.03	0.85**	0.14**	0.11**	0.00	-0.05	-0.06	-0.16**	-0.10**	
	% NCRU		0.13**	0.05	-0.03	-0.05	-0.03	0.54**	0.08*	0.08*	-0.01	-0.02	-0.05	-0.11**	-0.10*	
	% FMOL			0.62**	0.05	-0.03	-0.05	-0.03	-0.05	-0.03	0.24**	0.37**	-0.13**	0.01	0.13**	
	% NMOL				-0.06	-0.12**	-0.01	0.08*	-0.12**	-0.12**	0.23**	0.45**	-0.12**	0.10**	0.23**	
	% NOLI					-0.15**	-0.09*	-0.04	0.32**	0.23**	-0.20**	-0.13**	-0.15**	-0.32**	0.24**	
	% FPLE						0.43**	-0.04	-0.28**	-0.23**	0.18**	-0.23**	0.96**	0.30**	-0.34**	
	% NPLE							-0.02	-0.10**	-0.07	0.07	-0.09*	0.38**	0.11**	-0.17**	
	FCRU								0.07	0.07	-0.01	0.02	-0.02	-0.08*	-0.06	
	DOM2										0.89**	-0.77**	-0.14**	-0.32**	-0.96**	0.07
	DOM1											-0.86**	-0.08	-0.25**	-0.92**	0.06
	EVN												-0.03	0.19**	0.79**	-0.13**
	FMOL													-0.19**	0.16**	0.46**
	FPLE														0.34**	-0.30**
	SHDIV															-0.05

Cluster 2	Variables	% NNINS	% FTRI	% NTRI	FEPH	FEPT	FFAC	FBI	FINT	SMDIV	FTRI
% NINT		-0.40**	0.46**	0.14**	0.54**	0.66**	0.33**	-0.76**	0.72**	0.52**	0.57**
% NNINS			-0.37**	-0.34**	-0.39**	-0.44**	-0.30**	0.46**	-0.42**	-0.17**	-0.38**
% FTRI				0.54**	0.37**	0.68**	0.30**	-0.45**	0.66**	0.29**	0.85**
% NTRI					0.26**	0.37**	0.27**	-0.46**	0.36**	0.14**	0.45**
FEPH						0.87**	0.79**	-0.53**	0.73**	0.53**	0.59**
FEPT							0.72**	-0.62**	0.93**	0.55**	0.89**
FFAC								-0.42**	0.56**	0.43**	0.58**
FBI									-0.64**	-0.55**	-0.54**
FINT										0.54**	0.86**
SMDIV											0.43**

Cluster 3	Variables	% FEPH	% NEPH	% FEPT	% NEPT	% FFAC	% NFAC	% FINT	% FTOL	% NTOL	FAM
% NCHI		-0.21**	-0.50**	-0.31**	-0.61**	-0.08*	-0.51**	-0.33**	0.32**	0.75**	-0.37**
% FEPH			0.55**	0.78**	0.48**	0.33**	0.22**	0.51**	-0.57**	-0.38**	0.34**
% NEPH				0.61**	0.80**	0.17**	0.25**	0.55**	-0.54**	-0.69**	0.50**
% FEPT					0.70**	0.21**	0.25**	0.81**	-0.75**	-0.54**	0.52**
% NEPT						0.21**	0.54**	0.64**	-0.63**	-0.85**	0.57**
% FFAC							0.42**	-0.10*	-0.46**	-0.26**	0.38**
% NFAC								0.14**	-0.34**	-0.68**	0.30**
% FINT									-0.72**	-0.54**	0.50**
% FTOL										0.66**	-0.64**
% NTOL											-0.53**

* $P < 0.05$, ** $P < 0.01$.

or “0–10” (Stoddard *et al.*, 2009). We chose metric distributions across all sampling sites, because the field data reflected various environmental impacts resulting from minimal to maximal impact of pollution. The scores were expressed in continuous values ranging from 0 to 1 in this study, to be readily comparable with other biological and non-biological indicators.

MISS scores and other indices were arranged along with the ranks in Figure 7. MISS appeared to be most suitable for presenting the integrity in a linear shape, compared with other indices. In the case of “percent of EPT taxa (% EPT)”, linearity was observed in the high ranks, but showed numerous zero values at the lowest ranks (*i.e.*, no values given at the extremely polluted sites). This type of index would not be feasible for identifying the

severely polluted states. FBI also showed a linear gradient to the intermediate range from the highest rank; however, it abruptly ended at the lowest range of the ranks. Shannon diversity index appeared to be linear, but was not strongly discriminative at the lowest ranks (Fig. 7). The profiles of “KSI” appeared to be non-linear in the middle range of the ranks between 150 and 300. MISS and BMWP were consistently linear across a broad range. However, discrimination in BMWP was not clearly observed at the severely polluted sites, as illustrated by showing the abrupt end at the lowest ranks when the gradients at the lowest ranks were enlarged (Fig. 8). In MISS, however, a gradual decrease was consistently observed even in the range of the lowest ranks. Overall, the gradient of MISS appeared to be more distinctive

Table 4. Correlation coefficients between biological indices and water quality indicators.

Variables	Shannon index	% EPT	FBI	BMWP	KSI	BOD	T-N	NH ₃ -N	NO ₃ -N	T-P	PO ₄ -P
MISS	0.68**	0.82**	-0.78**	0.89**	-0.50**	-0.25**	-0.26**	-0.15**	-0.25**	-0.11**	-0.13**
Shannon index		0.51**	-0.61**	0.71**	-0.40**	-0.19**	-0.22**	-0.12**	-0.24**	-0.13**	-0.15**
% EPT			-0.53**	0.70**	-0.45**	-0.27**	-0.27**	-0.18**	-0.25**	-0.15**	-0.16**
FBI				-0.60**	0.43**	0.20**	0.18**	0.10*	0.18**	0.08*	0.08*
BMWP					-0.43**	-0.19**	-0.22**	-0.12**	-0.23**	-0.09*	-0.11**
KSI						0.45**	0.41**	0.34**	0.40**	0.32**	0.32**
BOD							0.70**	0.61**	0.61**	0.60**	0.60**
T-N								0.82**	0.94**	0.70**	0.71**
NH ₃ -N									0.67**	0.67**	0.66**
NO ₃ -N										0.62**	0.64**
T-P											0.98**

* $P < 0.05$, ** $P < 0.01$.

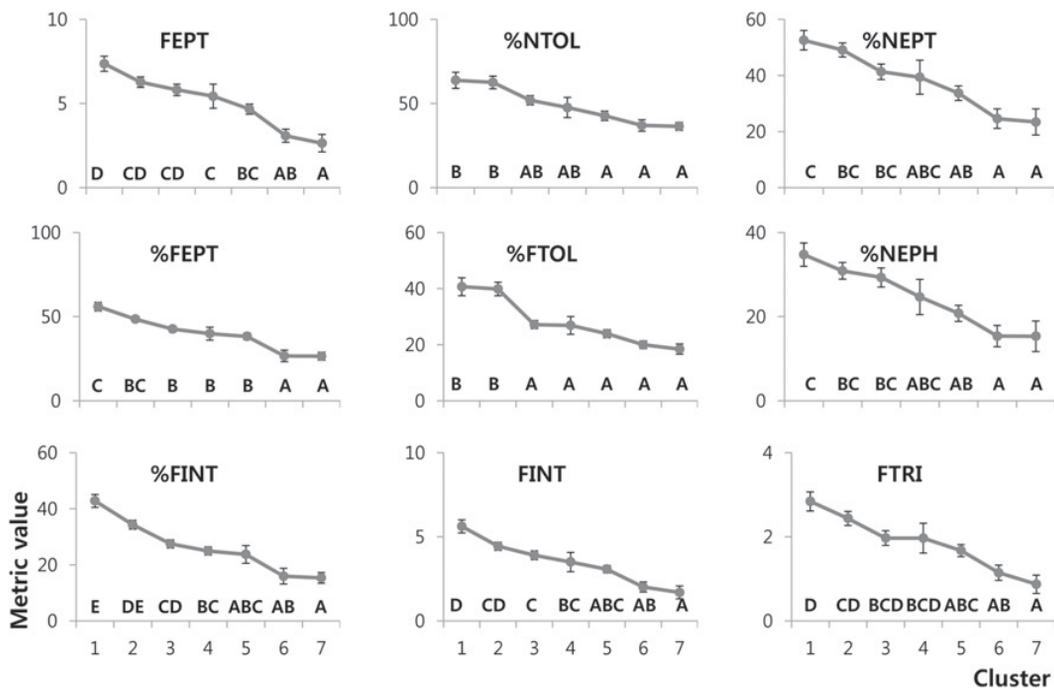


Fig. 6. Differences of the selected metrics at different clusters on the SOM trained with water quality indicators. Error bar indicates standard error. Different characters on each metric indicate significant differences among clusters based on the multiple comparison with Tukey HSD test ($P < 0.05$).

in a linear shape across a broad range of environmental impact.

MISS showed strong correlation with other indices in the range of 0.50–0.89 ($P < 0.01$), showing the highest value with BMWP and the lowest value with KSI, and negative correlation in lower ranges with chemical environmental factors including BOD in the range of -0.11 to -0.26 (Table 4). Among biological indices, other indices also were significantly correlated with each other. However, their relation was lower than that with MISS.

Multi-metric indicator

Taxonomic sufficiency at the family level was efficiently illustrated by producing a multi-metric indicator (MISS)

for presenting ecological integrity across different levels of environmental impact. The family-level metrics could overcome the problem of species level identification of benthic macroinvertebrates, which results in technical difficulties, increased time, and cost (Resh *et al.*, 1995; Waite *et al.*, 2004). Taxonomic resolution was compared using macroinvertebrates for bioassessment, and led to the possibility of using the family or genus levels for monitoring (Pond *et al.*, 2008; Trigo-Domínguez *et al.*, 2010). For the rapid assessment of streams using a macroinvertebrate community, efficient use of metrics at the family level was regarded as an adequate description of water quality (Barbour *et al.*, 1999). Green *et al.* (2000) used family-level data for investigating the impact of mining on the streams in a mountain. In this study, we confirmed that a multi-metric-based assessment on the family level was efficient

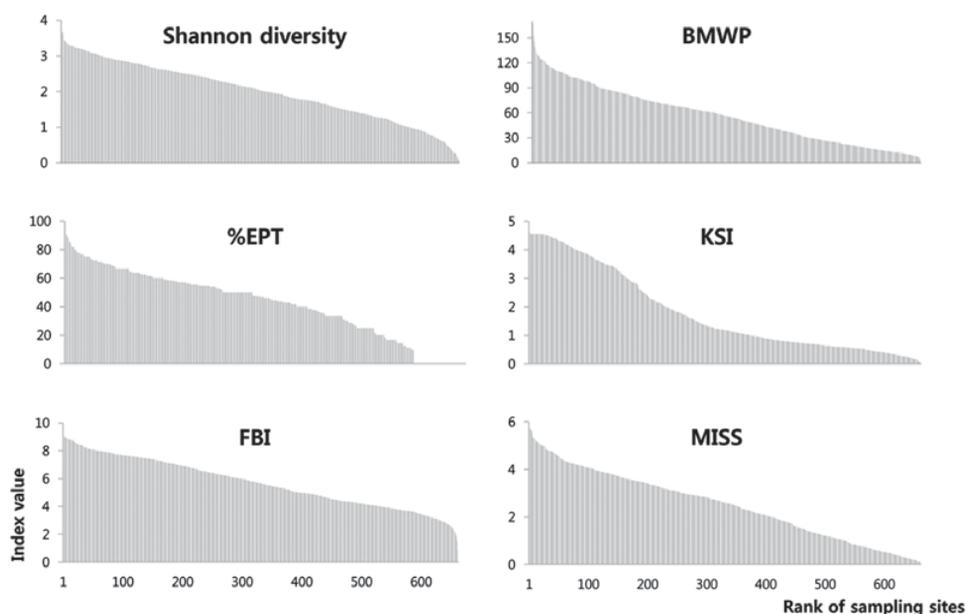


Fig. 7. Illustration of the values of MISS and biological indices according to rank.

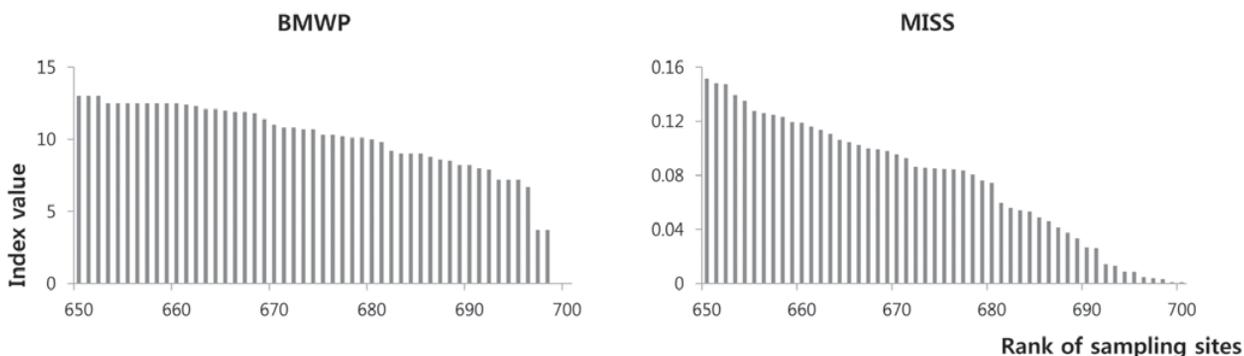


Fig. 8. Illustration of the values of BMWP and MISS in the lowest ranks of 50 sites.

in presenting ecological integrity in streams across different levels of environmental impact (Figs. 7 and 8, Table 4).

The efficiency of MISS was demonstrated for discriminating ecological integrity, especially in the lowest range of the ranks (*i.e.*, severely polluted sites). Although the BMWP values were linear in presenting water quality from intermediate to clean areas, discriminatory power was weaker at the severely polluted sites compared with MISS (Fig. 8). This result demonstrated feasibility of the multi-metric for presenting ecological integrity across a broad range of pollution. Since the multi-metric was composed of different single indices, each of the metrics would cover different sensitivity ranges in response to environmental impact. This type of diverse range covered by multi-metrics would help provide consistency in responding to disturbances, including severe pollution. The metric, % NTOL, was used in MISS, and this metric played the key role in differentiating the effect of severe pollution. The metric % NTOL includes densities of Chironomidae and Oligochaeta, which are highly tolerant to oxygen

depletion; these taxa effectively responded to severe pollution.

Since the proposed multi-metric was limitedly applied to field data in one nation, during one year in this study, more field tests are required to confirm the feasibility of the proposed metric for large-scale use across different levels of disturbance. In addition, the metric needs to be more closely compared with other indices in different regimes of ecosystem quality indicators such as saprobity and physico-chemical factors. For instance, the states of xenosaprobity and oligosaprobity could be compared with the multi-metric for expressing ecosystem health, and the water quality indices such as BOD or conductivity could be checked to be compatible with ecological integrity in different purposes of water usage (*e.g.*, drinking or biodiversity conservation). Guidelines may be needed to determine which indices should be recommended for the purpose of ecosystem management in the future. Since multi-metrics are based on biological properties, the metrics would be consequently more suitable for presenting ecological integrity regarding community

structure (e.g., maintenance of species diversity) or ecosystem function (e.g., maintaining high energy efficiency). In order to obtain in-depth views on ecosystem health assessment, however, more studies are required to elucidate the complex relationships between the metrics and the related ecological processes regarding community dynamics and ecosystem functioning through large-scale and long-term surveys in the future.

Conclusion

A multi-metric (MISS) based on the family-level classification of benthic macroinvertebrate communities was proposed as a suitable indicator for presenting ecological integrity across different levels of environmental impact. The screening processes aided by the SOM training were useful in extracting information from complex communities by estimating discrimination power, degree of redundancy, and responsiveness to stress. The proposed metric needs to be further tested regarding its relationship to community dynamics and ecosystem functioning through large-scale and long-term surveys in the future.

Acknowledgements. This study was financially supported by the Ministry of Environment and the National Institute of Environmental Research (Korea), and the results of this study form part of the “Survey and Evaluation of Aquatic Ecosystem Health in Korea, 2009”. The authors would like to thank all of survey members involved in the project for their help in sampling and analysis. We also thank reviewers for their constructive comments in improving the manuscript.

References

- Armitage P.D., Moss D., Wright J.F. and Furse M.T., 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.
- Barbour M.T., Stribling J.B. and Karr J.R. 1995. The multimetric approach for establishing biocriteria and measuring biological conditions. *In*: Davis W.S. and Simon T.P. (eds.), *Biological Assessment and Criteria, Tools for Water Resource Planning and Decision Making*, Lewis Publishers, FL.
- Barbour M.T., Gerritsen J., Griffith G.E., Frydenborg R., McCarron E., White J.S. and Bastian M.L., 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *J. N. Am. Benthol. Soc.*, 15, 185–211.
- Barbour M.T., Gerritsen J., Snyder B.D. and Stribling J.B., 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, Environmental Protection Agency, Washington, DC.
- Blocksom K.A., 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. *Environ. Manage.*, 31, 670–682.
- Blocksom K.A., Kurtenbach J.P., Klemm D.J., Fulk F.A. and Cormier S.M., 2002. Development and evaluation of the Lake Macroinvertebrate Integrity Index (LMII) for New Jersey lakes and reservoirs. *Environ. Monit. Assess.*, 77, 311–333.
- Böhmer J., Rawer-Jost C., Zenker A., Meier C., Feld C.K., Biss R. and Hering D., 2004. Assessing streams in Germany with benthic invertebrates: Development of a multimetric invertebrate based assessment system. *Limnologica*, 34, 416–432.
- Butcher J.T., Stewart P.M. and Simon T.P., 2003. A benthic community index for streams in the northern lakes and forests ecoregion. *Ecol. Indic.*, 3, 181–193.
- Carlisle D.M. and Clements W.H., 1999. Sensitivity and variability of metrics used in biological assessments of running waters. *Environ. Toxicol. Chem.*, 18, 285–291.
- Chon T.-S., Park Y.-S., Moon K.H. and Cha E.Y., 1996. Patternizing communities by using an artificial neural network. *Ecol. Model.*, 90, 69–78.
- Green J., Passmore M. and Childers H., 2000. A survey of the condition of streams in the primary region of mountaintop mining/valley fill coal mining. Appendix. *In*: Mountaintop Mining/valley Fills in Appalachia. Final programmatic environmental impact statement, US Environmental Protection Agency, Philadelphia, PA.
- Hargett E.G. and ZumBerge J.R., 2006. Redevelopment of the Wyoming Stream Integrity Index (WSII) for assessing the biological condition of wadeable streams in Wyoming, Wyoming Department of Environmental Quality, Water Quality Division, Cheyenne, WY.
- Hawkes H., 1979. Invertebrates as indicators of river water quality. *In*: James A. and Evison L. (eds.), *Biological Indicators of Water Quality*, John Wiley and Sons, Chichester.
- Hilsenhoff W.L., 1987. An improved biotic index of organic stream pollution. *Great Lake Entomol.*, 20, 31–40.
- Karr J.R. and Chu E.W., 1999. *Restoring life in running waters: better biological monitoring*, Island Press, Washington, DC, 206 p.
- Kerans B.L. and Karr J.R., 1994. A Benthic Index of Biotic Integrity (B-IBI) for Rivers of the Tennessee Valley. *Ecol. Appl.*, 4, 768–785.
- Klemm D.J., Blocksom K.A., Thoeny W.T., Fulk F.A., Herlihy A.T., Kaufmann P.R. and Cormier S.M., 2002. Methods development and use of macroinvertebrates as indicators of ecological conditions for streams in the Mid-Atlantic Highlands Region. *Environ. Monit. Assess.*, 78, 169–212.
- Klemm D.J., Blocksom K.A., Fulk F.A., Herlihy A.T., Hughes R.M., Kaufmann P.R., Peck D.V., Stoddard J.L., Thoeny W.T. and Griffith M.B., 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environ. Manage.*, 31, 656–669.
- Lenat D.R., 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. *J. N. Am. Benthol. Soc.*, 7, 222–233.
- Lenat D.R., 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water-quality ratings. *J. N. Am. Benthol. Soc.*, 12, 279–290.
- McNaughton S.J., 1967. Relationship among functional properties of California grassland. *Nature*, 216, 168–169.
- MOE/NIER, 2008. The survey and evaluation of aquatic ecosystem health in Korea, The Ministry of Environment/National Institute of Environmental Research, Incheon, Korea (in Korean with English summary).

- Ode P.R., Rehn A.C. and May J.T., 2005. A quantitative tool for assessing the integrity of southern coastal California streams. *Environ. Manage.*, 35, 493–504.
- Park Y.-S., Cereghino R., Compin A. and Lek S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecol. Model.*, 160, 265–280.
- Pielou E.C., 1975. *Ecological Diversity*, Wiley, New York.
- Pond G.J., Passmore M.E., Borsuk F.A., Reynolds L. and Rose C.J., 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *J. N. Am. Benthol. Soc.*, 27, 717–737.
- Purcell A.H., Bressler D.W., Paul M.J., Barbour M.T., Rankin E.T., Carter J.L. and Resh V.H., 2009. Assessment tools for urban catchments: developing biological indicators based on benthic macroinvertebrates. *J. Am. Water Resour. Assoc.*, 45, 306–319.
- Resh V.H., Norris R.H. and Barbour M.T., 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Aust. J. Ecol.*, 20, 108–121.
- Reynoldson T.B., Norris R.H., Resh V.H., Day K.E. and Rosenberg D.M., 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J. N. Am. Benthol. Soc.*, 16, 833–852.
- Rosenberg D.M. and Resh V., 1993. *Freshwater Biomonitoring and Benthic Macroinvertebrates*, Chapman and Hall, London, 488 p.
- Shannon C.E. and Weaver W., 1949. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.
- Simon T.P. and Lyons J., 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. In: Davis W.S. and Simon T.P. (eds.), *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*, Lewis Publishers, Chelsea, MI.
- Solimini A.G., Bazzanti M., Ruggiero A. and Carchini G., 2008. Developing a multimetric index of ecological integrity based on macroinvertebrates of mountain ponds in central Italy. *Hydrobiologia*, 597, 109–123.
- Song M.-Y., Hwang H.-J., Kwak I.-S., Ji C.W., Oh Y.-N., Youn B.J. and Chon T.-S., 2007. Self-organizing mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation. *Ecol. Model.*, 203, 18–25.
- Stoddard J.L., Herlihy A.T., Peck D.V., Hughes R.M., Whittier T.R. and Tarquinio E., 2009. A process for creating multimetric indices for large-scale aquatic surveys. *J. N. Am. Benthol. Soc.*, 27, 878–891.
- Taowu M., Qinghui H., Hai W., Zijian W., Chunxia W. and Shengbiao H., 2008. Selection of benthic macroinvertebrate-based multimetrics and preliminary establishment of bio-criteria for the bioassessment of the water quality of Taihu Lake, China. *Acta Ecol. Sin.*, 28, 1192–1200.
- The Mathworks Inc., 2001. *Matlab*. Version 6.1, The Mathworks Inc., Massachusetts.
- Trigal-Domínguez C., Fernández-Aláez C. and García-Criado F., 2010. Ecological assessment of highly heterogeneous systems: The importance of taxonomic sufficiency. *Limnologica*, 40, 208–214.
- Trigal C., Garcia-Criado F. and Fernandez-Alaez C., 2009. Towards a multimetric index for ecological assessment of Mediterranean flatland ponds: the use of macroinvertebrates as bioindicators. *Hydrobiologia*, 618, 109–123.
- Vesanto J., Himberg J., Alhoniemi E., Parhankangas J., Team S. and Oy L., 2000. *SOM Toolbox for Matlab*, Techn. Ber. Helsinki University of Technology.
- Waite I.R., Herlihy A.T., Larsen D.P., Urquhart N.S. and Klemm D.J., 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, USA. *Freshwater Biol.*, 49, 474–489.
- Walley W.J. and Hawkes H.A., 1996. A computer-based reappraisal of the Biological Monitoring Working Party scores using data from the 1990 river quality survey of England and Wales. *Water Res.*, 30, 2086–2094.
- Walley W.J. and Hawkes H.A., 1997. A computer-based development of the Biological Monitoring Working Party score system incorporating abundance rating, site type and indicator value. *Water Res.*, 31, 201–210.
- Ward J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 58, 236–244.
- Whittier T.R., Hughes R.M., Stoddard J.L., Lomnický G.A., Peck D.V. and Herlihy A.T., 2007. A structured approach to developing indices of biotic integrity: three examples from western USA streams and rivers. *Am. Fish. Soc.*, 136, 718–735.
- Won D.H., Jun T.C., Kwon S.J., Hwang S.J., Ahn K.G. and Lee J.K., 2006. Development of Korean Saprobic Index using benthic macroinvertebrate and its application to biological stream environment assessment. *J. Korean Soc. Water Qual.*, 22, 768–783.
- Wright J.F., Sutcliffe D.W. and Furse M.T., 2000. *Assessing the biological quality of fresh waters: RIVPACS and other techniques*, Freshwater Biological Association, Ambleside, 400 p.
- Zurada J., 1992. *Introduction to artificial neural systems*, West Publishing Co., NY, 683 p.