Relationships between three major stream assemblages and their environmental factors in multiple spatial scales

Mi-Jung Bae[1], Yongsu Kwon[1], Soon-Jin Hwang[2], Tae-Soo Chon[3], Hyung-Jae Yang[4], In-Sil Kwak[5], Jung-Ho Park[6], Soon-A Ham[7] and Young-Seuk Park[1]*

[1] Department of Biology, Kyung Hee University, Seoul 130-701, Republic of Korea
[2] Department of Environmental Science, Konkuk University, Seoul 143-701, Republic of Korea
[3] Department of Biological Science, Pusan National University, Busan 609-735, Republic of Korea
[4] Water Environment Research Department, The National Institute of Environmental Research, Incheon 404-170, Republic of Korea
[5] Division of Ocean and Fisheries, Chonman National University, Yosu 500-749, Republic of Korea
[6] Dong Lim P&D, Chuncheon 200-701, Republic of Korea
[7] Department of Biological Science, Chonman National University, Gwangju 500-757, Republic of Korea

Abstract – This study investigated the relationships of three major aquatic assemblages (diatom, macroinvertebrate, and fish) and environmental variables, including sub-basin, hydrology, land cover, and water quality variables on multiple scales. Samples were collected at 720 sampling sites on the Korean nationwide scale. Geological variables, including altitude and slope, showed a strong positive correlation with proportions of forest in land cover types and cobbles in substrates, while they were negatively correlated with water quality variables, including conductivity and total phosphorus. Considering the concordance of the different assemblages, species richness of fish and macroinvertebrates displayed significant correlation, and diatoms were significantly correlated with fish. However, diatoms did not show significant correlation with macroinvertebrates. Altitude and slope showed significant correlation with all biological variables of the three assemblages. Macroinvertebrates and fish showed positive relations with large substrate sizes. Indices of diatoms and macroinvertebrates well reflected the perturbation of water quality variables. However, fish indices showed a relatively low association with water quality variables, compared with those of diatoms and macroinvertebrates. These patterns were also confirmed by the ordination and prediction of biological indices with environmental variables through the learning process of a self-organizing map as well as random forest. Overall, our study supports the concept of multi-scale habitat filters and functional organization in streams, and is consistent with the recommended use of multiple biological indices with more than one assemblage for the assessment of the biotic integrity of aquatic ecosystems.

Key words: Multimetric / concordance of taxa / biological indices / self-organizing map / random forest

Introduction

Different biological assemblages display different sensitivities to diverse stressors on different scales. The United States Environmental Protection Agency (US EPA, 2002) uses three different biological assemblages (periphytons, macroinvertebrates, and fish) in the biomonitoring and assessment of water quality and stream biological integrity. The European Water Framework Directive (European Commission, 2000) also advocates the use of these organism groups in assessment of the ecological integrity of stream ecosystems. As a result of differences in life history and biogeography among these three assemblages, they may differ in physical and chemical tolerances (Townsend and Hildrew, 1994; Griffith et al., 2005). For example, due to their short generation time (Rott, 1991), diatoms may show a rapid, direct, and sensitive response to many physical, chemical, and biological changes, compared with other assemblages, including macroinvertebrates and fish (Stevenson and Pan, 1999). Thus, diatoms are typically used for assessment of nutrient enrichment, including trophic state, saprobity, and oxygen requirement (Hering et al., 2006). By linking producers, top carnivores, and decomposers in aquatic communities, macroinvertebrates play a key role as consumers in the food trophic structure (Allan and Castillo, 2007).

*Corresponding author: parkys@khu.ac.kr

Macroinvertebrates are sedentary and have an intermediate lifespan ranging from months to a few years; therefore, they have been widely used for the ecological assessment of water quality in an integrative and continuous manner (Park et al., 2007). Fish, which are the longest-lived organisms in aquatic environments as well as being migratory organisms, can be found everywhere in aquatic ecosystems and play a major role in aquatic food-webs as a carrier of energy from lower to higher trophic levels (Beyer, 1996). Thus, fish can be valuable indicators for the assessment of long-term and larger spatial scale events, such as river continuity disruption, alterations in catchment level, land cover, and land use (Johnson et al., 2007).

Biological indices for diatoms, macroinvertebrates, and fish assemblages can be effective for use in monitoring of stream enrichment; however, little is known with regard to the value of the different assemblages (Justus et al., 2010) for the detection of variations in environmental factors on various scales. Therefore, the evaluations of ecological integrity based on more than one assemblage for the diagnosis of environmental stressors on multiple scales have increased. For example, Griffith et al. (2005) assessed the relative sensitivity of assessments using community metrics for periphyton, macroinvertebrates, and fish assemblages based on environmental factors of chemical and physical habitat. Hering et al. (2006) showed different responses among four organism groups (periphytic diatoms, macrophytes, benthic macroinvertebrates, and fish) in terms of physical–chemical characteristics, hydromorphology, and land use in streams of European countries. Justus et al. (2010) reported higher correlations of nutrient concentrations with the algal index than with the macroinvertebrate and fish indices.

Distribution and abundance of organisms are influenced by various factors, including biogeographical constraints, as well as landscape features, such as topography, geology, and climate at multiple spatial scales (Infante et al., 2009). Poff (1997) proposed multi-scale habitat filters and functional organization in streams, specifying a set of four habitat levels (watershed, reach, channel unit, and microhabitat). However, slightly different numbers of levels and diversity of elements within levels have been described by other authors (Gregory et al., 1991; Townsend, 1996). Landscape features affect small-scale factors, such as microhabitat conditions and water quality, which are important to the distribution and abundance of organisms. Environmental conditions can be viewed as constituting filters through which species in the regional species pool must pass to be potentially present at a given locale (Tonn et al., 1990; Poff, 1997).

Although a great deal of research using various assemblages has been conducted for the evaluation of ecosystem health, only minimal information has been available for full evaluation of environmental and biological quality on the national scale, particularly in Korea (e.g., Park et al., 2007). In addition, relatively few studies have evaluated the response of multiple biological assemblages in biomonitoring studies, and even fewer studies have examined the relationship and intercorrelation of

different assemblages' measures and their response to stress on multiple spatial scales (Hering et al., 2006; Flinders et al., 2008). In this sense, our study had three aims: to evaluate the relationships between diatom, macroinvertbrate, and fish indices as descriptive measures in streams; to examine the relationships between biological indices and environmental variables in multi-levels, such as sub-basin, hydrology, land cover, substrate, and water quality variables; and to evaluate important environmental variables influencing biological indices. Our study may be helpful in the effort to better understand the present ecosystem status of streams in Korea and may provide information needed for the establishment of standards for the ecological health and for the restoration of damaged stream ecosystems.

Materials and methods

Ecological data

Data were obtained through the National Aquatic Ecological Monitoring Program (NAEMP) operated by the Ministry of Environment and National Institute of Environmental Research, Korea. Three assemblages (diatom, macroinvertebrate, and fish) were extensively surveyed at 720 sampling sites. In order to avoid community variability compounded with the season or with heterogeneity in sampling sites, all samples were collected during normal flow periods, especially from April to May in 2009, according to the guidelines of the "National Surveys for Stream Ecosystem Health" in Korea (MOE/NIER, 2008).

Diatom samples were collected from three arbitrarily selected cobbles at each sampling site. Using a toothbrush, diatoms were scraped from a 10 cm^2 area of the upper part of each cobble substratum. Sampled diatoms were preserved using formalin or Lugol's iodine solution. From each sample, at least 300 diatom valves were identified and counted at $1000\times$ magnification. Relative abundances of the diatom species (in percentage) were used in this study. Details of the sampling protocol for diatoms are described in MOE/NIER (2008).

Macroinvertebrates were sampled with a Surber net (30×30 cm, 1 mm mesh size) in the riffle zones of streams. Samples were preserved with formalin or 95% ethanol (70% final concentration). In the laboratory, invertebrate specimens were sorted and identified, mostly to the species level, and the numbers of specimens were counted under microscopes. Details of the sampling protocol for macroinvertebrates are described in MOE/NIER (2008).

Fish were collected from all types of habitats, including riffle, run, and pool areas based on the catch per unit effort. Stream segments of approximately 200 m were sampled at each site for 50 min. Fish samples were collected with two types of sampling equipment: a casting net (5 mm mesh size) and a kick net (4 mm mesh size). Most of the captured fish were identified to the species level in the field. Among the collected specimens, some

Table 1. Characteristics of environmental variables of the study sites.

Variables	Abbreviation	N	Mean (S.E.*)	Range	Transformation
Sub-basin					
Altitude (m)		720	106.11 (4.70)	0.00–721.00	$x' = \ln(x + 1)$**
Slope (%)		720	13.19 (0.66)	1.00–91.25	
Distance from the source (km)		720	50.47 (3.00)	0.21–49.89	$x' = \ln(x + 1)$
Stream order		720	5.25 (0.06)	1.00–9.00	
Land cover					
Urban (%)		719	21.40 (1.05)	0.00–100.00	
Paddy field (%)		719	32.13 (1.18)	0.00–100.00	
Dry field (%)		719	14.43 (0.65)	0.00–99.94	
Forest (%)		719	32.03 (1.19)	0.00–100.00	
Hydrology					
Stream width (m)	Width	649	104.50 (3.10)	1.00–350.00	
Current velocity (cm.s^{-1})	Velocity	717	33.84 (1.16)	0.00–137.70	
Substrate					
Silt (%)		716	5.02 (0.48)	0.00–80.00	
Fine sand (%)		716	9.27 (0.59)	0.00–90.00	
Coarse sand (%)		716	20.83 (0.82)	0.00–100.00	
Small gravel (%)		716	18.08 (0.49)	0.00–84.21	
Gravel (%)		716	21.73 (0.58)	0.00–80.00	
Cobble (%)		716	21.87 (0.79)	0.00–100.00	
Boulder (%)		716	3.21 (0.34)	0.00–90.00	
Water quality					
Dissolved oxygen (mg.L^{-1})	DO	701	10.26 (0.09)	4.04–16.54	
pH		718	8.12 (0.03)	6.24–10.60	
Electric conductivity (uS.cm^{-1})	Conductivity	667	248.98 (5.68)	1.00–735.00	
Turbidity (NTU)		563	6.27 (0.25)	0.00–29.30	
Biological oxygen demand (mg.L^{-1})	BOD	669	2.41 (0.06)	0.30–7.20	
Total nitrogen (mg.L^{-1})	TN	682	2.53 (0.06)	0.29–7.64	
Ammonia nitrogen (mg.L^{-1})	NH$_3$-N	603	0.03 (0.00)	0.00–0.27	
Nitrate nitrogen (mg.L^{-1})	NO$_3$-N	685	1.92 (0.05)	0.00–5.63	
Total phosphorus (mg.L^{-1})	TP	642	0.06 (0.00)	0.00–0.29	
Phosphate-phosphorus (mg.L^{-1})	PO$_4$-P	632	0.03 (0.00)	0.00–0.19	

*S.E.: standard error.

**x: a raw value of environmental variable, x': transformed value.

species requiring detailed identification and observation were fixed with 10% formalin solution and transported to the laboratory. Details of the sampling protocol for fish are described in MOE/NIER (2008).

Among a number of sub-basin, hydrology, land cover, physical, and water quality variables measured in the program, 27 environmental variables were used in this study (Table 1). Variables such as altitude, slope, distance from the source, and stream order were extracted from a digital map with ArcView (ver. 3.2). In addition, proportions (%) of land cover types, including urban, paddy fields, dry fields, and forests were also extracted at each transect within each riparian zone with 200 m wide buffers and 1 km length from the sampling sites on the digital map. Hydrological variables, substrates, and some water quality variables, including dissolved oxygen (DO), pH, electric conductivity, and turbidity were measured *in situ*, whereas other chemical variables, including biological oxygen demand (BOD), total nitrogen (TN), ammonia nitrogen (NH$_3$-N), nitrate nitrogen (NO$_3$-N), total phosphorus (TP), and phosphate-phosphorus (PO$_4$-P) were

measured in the laboratory after water sampling from the field (APHA, 2005).

Multimetric approach

Three biological indices in three different assemblage groups were selected for the determination of the relationships between biological indices and environmental factors (Table 2). Species richness (the actual number of species in a given area) in each assemblage is the basic step of community study and is fundamental for dealing with biodiversity as well as ecosystem management. The trophic diatom index (TDI) is an index for the monitoring of trophic status in freshwater ecosystems based on diatom composition (Kelly and Whitton, 1995). TDI is calculated using relative abundance in the sample, pollution sensitivity values, and indicator values in each taxon. The index ranges from 0 (oligotrohpic status) to 100 (polytrophic status). Diatom index for organic pollution (DAIpo; Watanabe *et al.*, 1986) has been widely used for

Table 2. Metrics included in three assemblages (diatom, macroinvertebrate, and fish).

Assemblages	Metric description	Abbreviation	*N*	Range	Expected response* (reference)
Diatom	Number of diatom taxa	D_SR	720	≥ 0	Decrease (Barbour *et al.*, 1999)
	Trophic diatom index (Kelly and Whitton, 1995)	TDI	720	0–100	Increase (Tang *et al.*, 2006)
	Diatom index for organic pollution (Watanabe *et al.*, 1986)	DAIpo	720	0–100	Decrease (Watanabe *et al.*, 1986)
Macro-invertebrate	Number of macroinvertebrate species	M_SR	712	≥ 0	Decrease (Barbour *et al.*, 1999)
	Average score per taxon (Armitage *et al.*, 1983)	ASPT	707	≥ 0	Decrease (Armitage *et al.*, 1983)
	Korean saprobic index (Won *et al.*, 2006)	KSI	701	0–5	Decrease (Won *et al.*, 2006)
Fish	Number of fish species	F_SR	699	≥ 0	Decrease (Barbour *et al.*, 1999)
	Relative abundance of sensitive species (MOE/NIER, 2008)	%_Sensi	699	0–100	Decrease (Barbour *et al.*, 1999)
	Korean index of biotic integrity (MOE/NIER, 2008)	K-IBIF	699	8–40	Decrease (Karr, 1981)

*Expected response to increasing disturbance.

the determination of organic pollution (Duong *et al.*, 2007). DAIpo is calculated based on the level of tolerance to organic pollution with three groups of diatoms, including saproxenous, saprophilous, and indifferent taxa. DAIpo values range from 0 (polytrophic) to 100 (oligotrophic status).

Average score per taxon (ASPT) in macroinvertebrates represents the average tolerance score of all taxa within the community, and it is calculated by dividing the revised biological monitoring working party score (revised BMWP; Walley and Hawkes, 1997) by the number of families presented in the sample. Although the revised BMWP has been widely used for measurement of water quality, it is dependent on the number of taxa, which can vary according to the sampling efforts and seasonal changes. Therefore, in this study, we applied it to ASPT. A higher ASPT score was considered to reflect better water quality. The Korean saprobic index (KSI; Won *et al.*, 2006), which ranges from 0 (oligotrophic status) to 5 (polytrophic status), was developed for evaluation of the saprobicity of water bodies using DIN 38410 (1990), based on the formula of Zelinka and Marvan (1961).

Relative abundance (%) of sensitive species in fish (%_Sensi) indicates various chemical and physical perturbations. Sensitive species are the first to disappear after a disturbance. In addition, %_Sensi tends to increase with small- and medium-sized streams, but is unaffected by large rivers (US EPA, 2002). Last, the Korean index of biological integrity using fish (K-IBIF) model (An and Lee, 2006; MOE/NIER, 2008) was developed for the evaluation of water quality and is based principally on Karr's IBI (Karr, 1981). Originally, Karr's model consisted of 12 biological metrics. However, K-IBIF used in this study was modified to fit the Korean environment, river structure, and species, and finally included eight metrics. K-IBIF, which ranges from 8 (polytrophic status) to 40 (oligotrophic status), is sensitive to a wide variety of changes in environmental conditions.

Data analysis

Spearman rank correlation coefficients were calculated to determine the relationships between the biological indices of three different assemblages, and environmental variables and biological indices.

For the classification of sampling sites based on similarities of the environmental conditions, we used a self-organizing map (SOM) (Kohonen, 2001). SOM is an unsupervised learning algorithm of artificial neural networks and approximates the probability density function of the input data (Kohonen, 2001). 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.

During the learning process of SOM, the input data (27 environmental variables in this study) were initially subjected to the network. The number of output neurons was set to 135 (15×9) in a two-dimensional hexagonal lattice based on operator experience and a preliminary study. Subsequently, the weights of the network were trained for a given dataset. Each node of the output layer computes the summed distance between weight vectors and input vectors. The output nodes are considered as virtual units for representation of 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 according to the weight vectors they are closest to. For the training SOM, we used the functions provided in the SOM toolbox (http://www.cis.hut.fi/projects/somtoolbox/) in Matlab (ver. 6.1).

After the training SOM, a unified distance matrix algorithm (U-matrix; Ultsch, 1993) and a hierarchical cluster analysis using Ward's linkage method based on the Euclidian distance were applied to the weights of the neurons in the SOM for further clustering of SOM units (Jain and Dubes, 1988; Park *et al.*, 2003). Then, using PC-ORD (ver. 5.3), multi-response permutation procedures (MRPP) were conducted to determine whether or not there was a significant difference among the clusters.
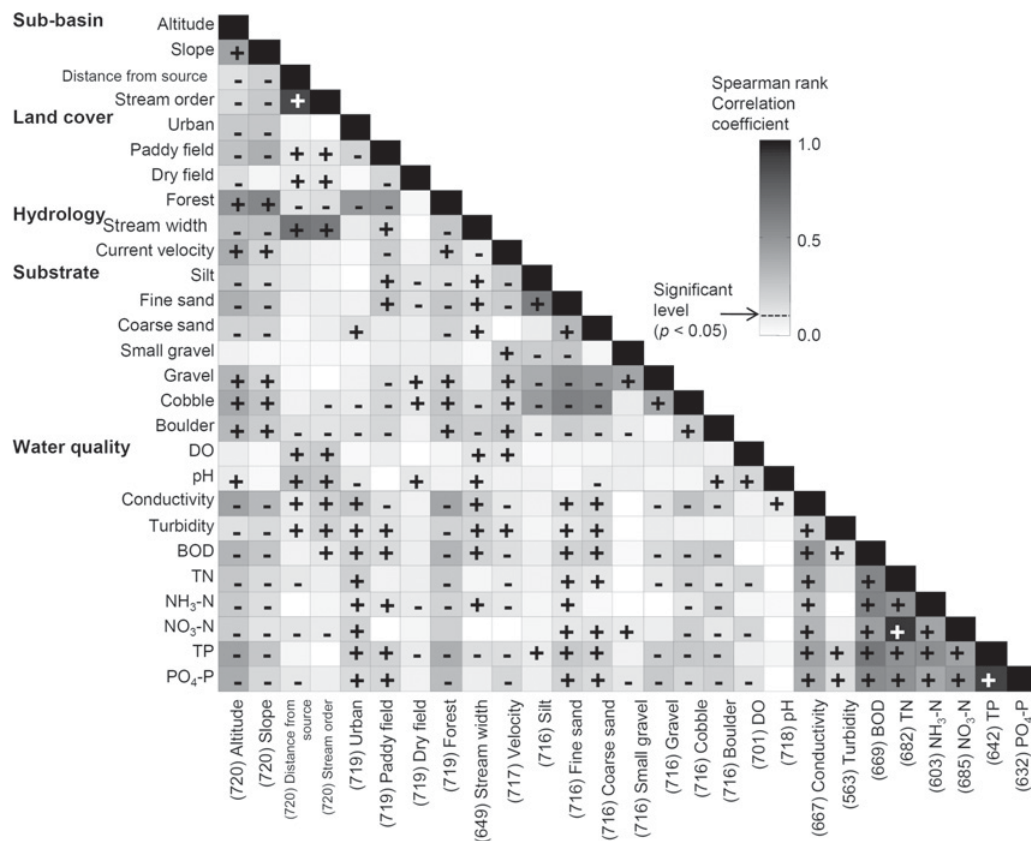
Fig. 1. Spearman rank correlation between 27 environmental variables. $+$, positive correlation ($P < 0.05$); $-$, negative correlation ($P < 0.05$); (n), number of samples.

The Kruskal–Wallis (K–W) test was conducted for the evaluation of the differences in environmental variables and community indices in different clusters defined in the SOM, and Dunn's multiple comparison tests were performed if significant differences were observed in the K–W test ($P < 0.05$).

Random forest was used for the prediction of both clusters defined in the SOM and biological indices with environmental variables, and for the evaluation of the contribution of each environmental variable in the determination of clusters and biological indices. Random forest is a non-parametric method for the prediction and assessment of the relationship among a large number of potential predictor variables and a response variable (Breiman, 2001). For measurement of the feature importance, we used mean decrease Gini (MDG), which quantifies the importance of a variable by summing all decrease in Gini impurity due to a given variable, when the values of the variable are randomly permuted, compared to the original observations. Values of MDG for each environmental variable were rescaled for each cluster or biological indices in the range of 0–100 in order to represent the relative importance of each environmental variable. The analysis was carried out with R-package *randomForest* (Liaw and Wiener, 2002), available at http://cran.r-project.org.

Prior to the analysis of environmental variables, variables showing high variations were transformed by natural logarithm prior to the process of quantitative data analysis (Table 1). In addition, to avoid the problem of logarithm zeros, the number one was added. STATISTICA (StatSoft, 2004) was used for statistical testing.

Results

Relationships between environmental variables

Considering the correlations among 27 environmental variables, altitude showed positive correlation with slope, forest, and cobbles in substrates ($r = 0.43$, 0.49, and 0.40, respectively, $P < 0.01$), while altitude was negatively correlated with water quality variables, including conductivity and TP ($r = -0.43$ and -0.44, respectively, $P < 0.01$), except for DO and pH (Fig. 1). Similarly, the proportion of forest in land covers showed a negative correlation with water quality variables, including conductivity and TP ($r = -0.41$ and -0.36, respectively, $P < 0.01$), while exhibiting a positive correlation with slope ($r = 0.57$, $P < 0.01$). Correlation between TN and NO$_3$-N was the highest ($r = 0.93$, $P < 0.01$) of the water quality variables, followed by TP and PO$_4$-P ($r = 0.90$, $P < 0.01$), BOD and TP ($r = 0.65$, $P < 0.01$), and BOD and NH$_3$-N ($r = 0.58$, $P < 0.01$).
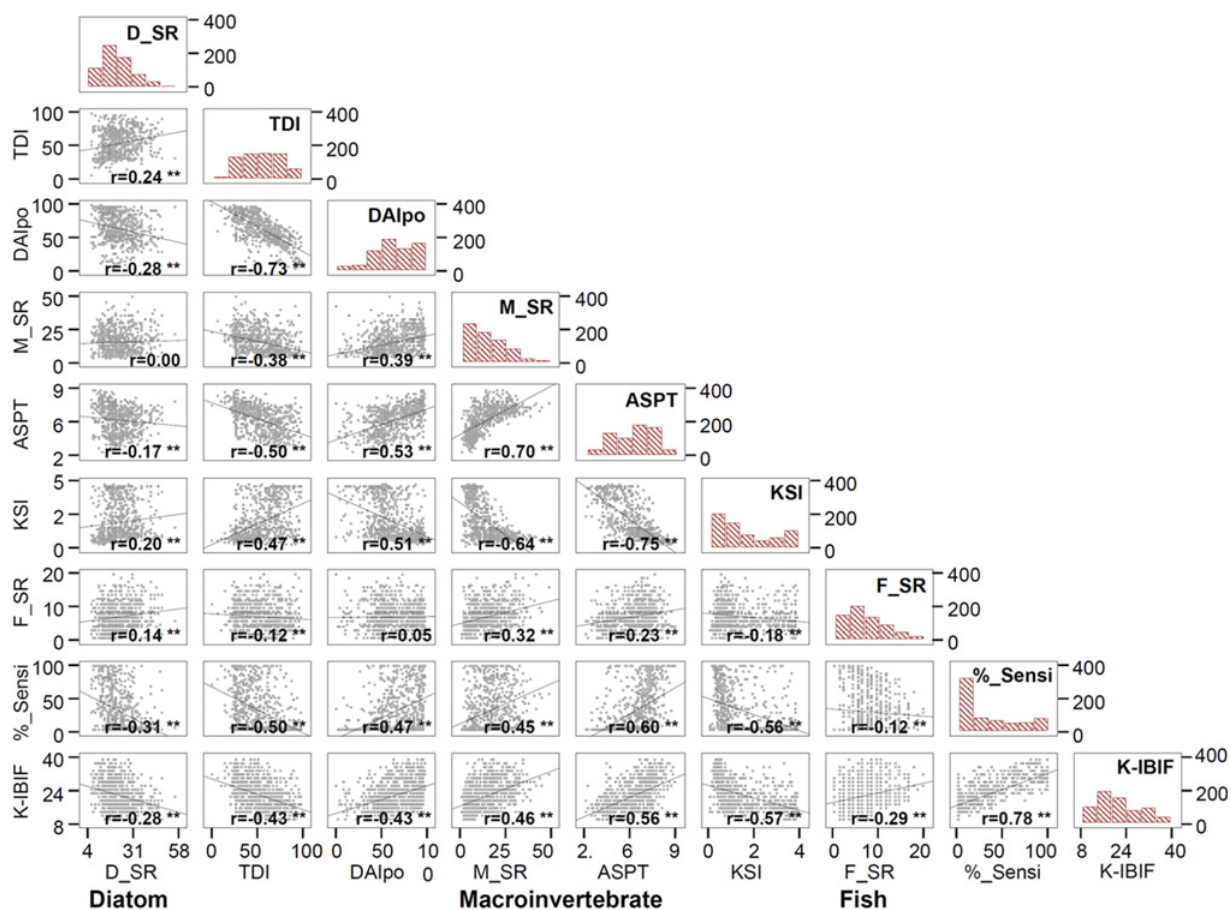
Fig. 2. Spearman rank correlation between biological indices in the three different assemblages (diatom, macroinvertebrate, and fish). Abbreviations of the variables are shown in Table 2. **$P < 0.01$.

Relationships between environmental and biological variables

Species richness of fish showed strong correlation with that of macroinvertebrates ($r = 0.32$, $P < 0.01$), whereas species richness of diatoms did not show either significant or weak correlation with that of macroinvertebrates ($r = 0.00$, $P > 0.05$) and fish ($r = 0.14$, $P < 0.01$), respectively (Fig. 2). Species richness of macroinvertebrates was highly correlated with ASPT ($r = 0.70$, $P < 0.01$) and KSI ($r = -0.64$, $P < 0.01$), whereas species richness of diatoms was weakly correlated with both TDI and DAIpo ($r = 0.24$ and -0.28, $P < 0.01$, respectively). Species richness of fish also displayed a low correlation coefficient with K-IBIF and %_Sensi ($r = 0.29$ and 0.12, $P < 0.01$, respectively). The correlation coefficient between indices within the same assemblages (*i.e.*, TDI and DAIpo in diatoms, ASPT and KSI in macroinvertebrates, and K-IBIF and % sensitive species abundance of fish), except for species richness, was significantly high ($r = -0.73$, -0.78, and 0.78, $P < 0.01$). Meanwhile, considering the relationships among biological indices in different assemblage groups, the highest correlation was observed between ASPT in macroinvertebrates and %_Sensi in fish ($r = 0.60$, $P < 0.01$), followed by KSI and K-IBIF ($r = -0.57$, $P < 0.01$), and KSI and

%_Sensi ($r = -0.56$, $P < 0.01$). Correlation coefficients of biological indices, except for species richness, between diatoms and fish were relatively lower than those of other cases ($r = |0.43–0.50|$, $P < 0.01$), although they were statistically significant (Fig. 2).

Altitude and slope on the sub-basin scale showed strong correlation with all biological variables of the three assemblages used in the analysis, although there was high variation (Table 3). Distance from the source and stream order was mostly correlated with fish indices, indicating high species richness of fish in larger streams ($r = 0.35$ and 0.32 respectively, $P < 0.01$). Indices of all of the assemblage groups showed a relatively high correlation with forest in land cover. In high ratio of forest, species richness of macroinvertebrates increased ($r = 0.33$, $P < 0.01$), while that of diatoms decreased ($r = -0.18$, $P < 0.01$). Species richness of fish did not show significant correlation with forest. Urban and paddy field showed positive correlation with TDI and KSI, while showing negative correlation with other indices, including DAIpo, ASPT, %_Sensi, and K-IBIF. Macroinvertebrates and fish displayed positive relationships with large substrate sizes, including gravel, cobble, and boulders. In particular, cobble showed the strongest correlation with %_Sensi and K-IBIF ($r = 0.40$ for both, $P < 0.01$), and ASPT and KSI ($r = 0.37$ and

Table 3. Spearman rank correlation between biological indices and environmental factors. Abbreviations of biological indices are shown in Table 2.

Environmental variables	Diatom			Macroinvertebrate			Fish		
	D_SR	TDI	DAIpo	M_SR	ASPT	KSI	F_SR	%_Sensi	K-IBIF
Sub-basin									
Altitude	− 0.26**	− 0.43**	0.45**	0.48**	0.58**	− 0.54**	0.15**	0.56**	0.54**
Slope	− 0.15**	− 0.26**	0.26**	0.25**	0.38**	− 0.32**	− 0.03	0.39**	0.39**
Distance from the source	0.11**	0.00	0.04	0.04	0.09*	0.00	0.35**	− 0.28**	− 0.23**
Stream order	0.16**	0.07	− 0.02	0.02	0.03	0.03	0.32**	− 0.33**	− 0.27**
Land cover									
Urban	0.11**	0.23**	− 0.19**	− 0.26**	− 0.29**	0.28**	− 0.09*	− 0.22**	− 0.23**
Paddy field	0.21**	0.17**	− 0.20**	− 0.06	− 0.22**	0.22**	0.15**	− 0.24**	− 0.23**
Dry field	− 0.02	− 0.09*	0.15**	0.12**	0.16**	− 0.13**	0.04	0.07	0.12**
Forest	− 0.18**	− 0.37**	0.34**	0.33**	0.47**	− 0.43**	0.04	0.43**	0.45**
Hydrology									
Stream width	0.21**	0.12**	− 0.14**	− 0.10*	− 0.10*	0.15**	0.30**	− 0.40**	− 0.32**
Current velocity	− 0.08*	− 0.08*	0.19**	0.40**	0.40**	− 0.37**	0.12**	0.17**	0.28**
Substrate									
Silt	0.03	0.10**	− 0.07*	− 0.24**	− 0.20**	0.17**	− 0.13**	− 0.21**	− 0.21**
Fine sand	0.09*	0.23**	− 0.21**	− 0.34**	− 0.31**	0.30**	− 0.11**	− 0.34**	− 0.37**
Coarse sand	0.09*	0.25**	− 0.18**	− 0.23**	− 0.23**	0.28**	− 0.09*	− 0.25**	− 0.28**
Small gravel	0.04	− 0.06	0.07*	0.16**	0.10**	− 0.05	0.03	0.02	0.07
Gravel	− 0.04	− 0.23**	0.24**	0.37**	0.30**	− 0.26**	0.18**	0.21**	0.27**
Cobble	− 0.14**	− 0.31**	0.29**	0.38**	0.37**	− 0.38**	0.15**	0.40**	0.40**
Boulder	− 0.17**	− 0.17**	0.24**	0.23**	0.30**	− 0.33**	− 0.01	0.28**	0.30**
Water quality									
DO	0.24**	− 0.05	0.08	0.24**	0.13**	− 0.14**	0.14**	− 0.02	− 0.01
pH	− 0.11**	0.07	0.01	0.03	0.07	− 0.12**	0.11**	− 0.01	0.06
Conductivity	0.31**	0.41**	− 0.44**	− 0.28**	− 0.48**	0.47**	− 0.14**	− 0.37**	− 0.43**
Turbidity	0.12**	0.20**	− 0.04	− 0.05**	− 0.12**	0.08	0.01	− 0.19**	− 0.15**
BOD	0.31**	0.43**	− 0.45**	− 0.29**	− 0.48**	0.47**	0.02	− 0.38**	− 0.44**
TN	0.18**	0.34**	− 0.35**	− 0.34**	− 0.40**	0.37**	− 0.16**	− 0.24**	− 0.30**
NH_3-N	0.26**	0.33**	− 0.47**	− 0.24**	− 0.42**	0.38**	0.02	− 0.35**	− 0.34**
NO_3-N	0.13**	0.33**	− 0.30**	− 0.31**	− 0.35**	0.32**	− 0.21**	− 0.21**	− 0.26**
TP	0.23**	0.49**	− 0.49**	− 0.38**	− 0.56**	0.52**	− 0.05	− 0.39**	− 0.38**
PO_4-P	0.14**	0.42**	− 0.42**	− 0.37**	− 0.50**	0.46**	− 0.13**	− 0.31**	− 0.34**

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

− 0.38, respectively, $P < 0.01$). In addition, the indices of diatom and macroinvertebrate well reflected the perturbation of water quality variables, displaying high correlation with nutrient concentration, showing the highest values in TP ($r = |0.49|$ with TDI and DAIpo in diatoms, and $r = − 0.56$ with ASPT in macroinvertebrates) (Table 3). However, correlation coefficients between fish indices and water quality variables were relatively lower than those of diatoms and macroinvertebrates, although fish were also influenced by water quality.

Classification of sampling sites

SOM classified 720 sampling sites based on the similarities of 27 environmental variables through the learning process of SOM (quantization error = 0.740 and topographic error = 0.039), and the SOM output units were classified into seven clusters (1–7) based on a hierarchical cluster analysis (Fig. 3). MRPP showed significant differences of environmental conditions among these clusters ($A = 0.09$, $P < 0.01$). Classification of sampling sites was highly related to their geographic location (Fig. 4). For example, sampling sites in cluster 1 were mainly in the Han River watershed, located in mountainous areas, sampling sites in cluster 4 were mainly from the main channels of the Han River, Nakdong River, and Seomjin River, sampling sites in cluster 6 were from urban areas along the Han River watershed, and sampling sites in cluster 7 were from the Han River, Geum River, and Yeongsan River watersheds in the western part of the lower-altitude Korean peninsula.

Values obtained from the weight vectors for the 27 environmental variables in the trained SOM were re-scaled, ranging from 0 to 1 (Fig. 5), and clusters were arranged in the gradient of altitude from high values to low values. Therefore, altitude and slope were the highest in cluster 1, whereas they were the lowest in cluster 7. In cluster 2, altitude was relatively high, but slope was low. Altitude had strong influences on other environmental
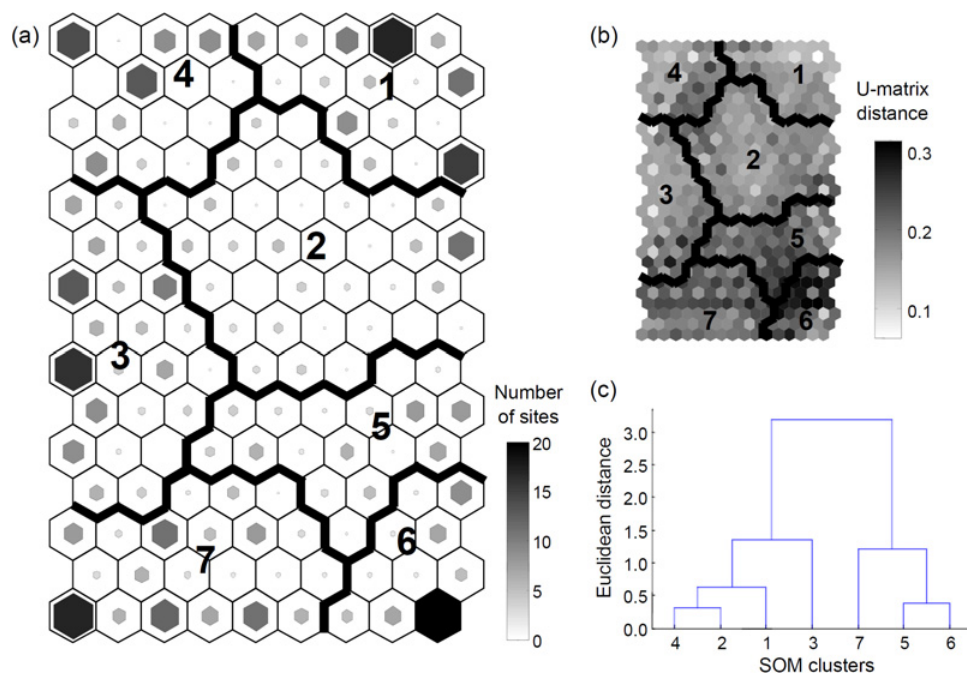
Fig. 3. (a) Classification of 720 sampling sites based on the SOM trained with 27 environmental variables measured in 2009. Gray-scale hexagons in each SOM unit represent the number of samples assigned in each SOM unit. (b) U-matrix distance and (c) dendrogram of a hierarchical cluster analysis of the SOM units using the Ward linkage method with Euclidean distance measure.
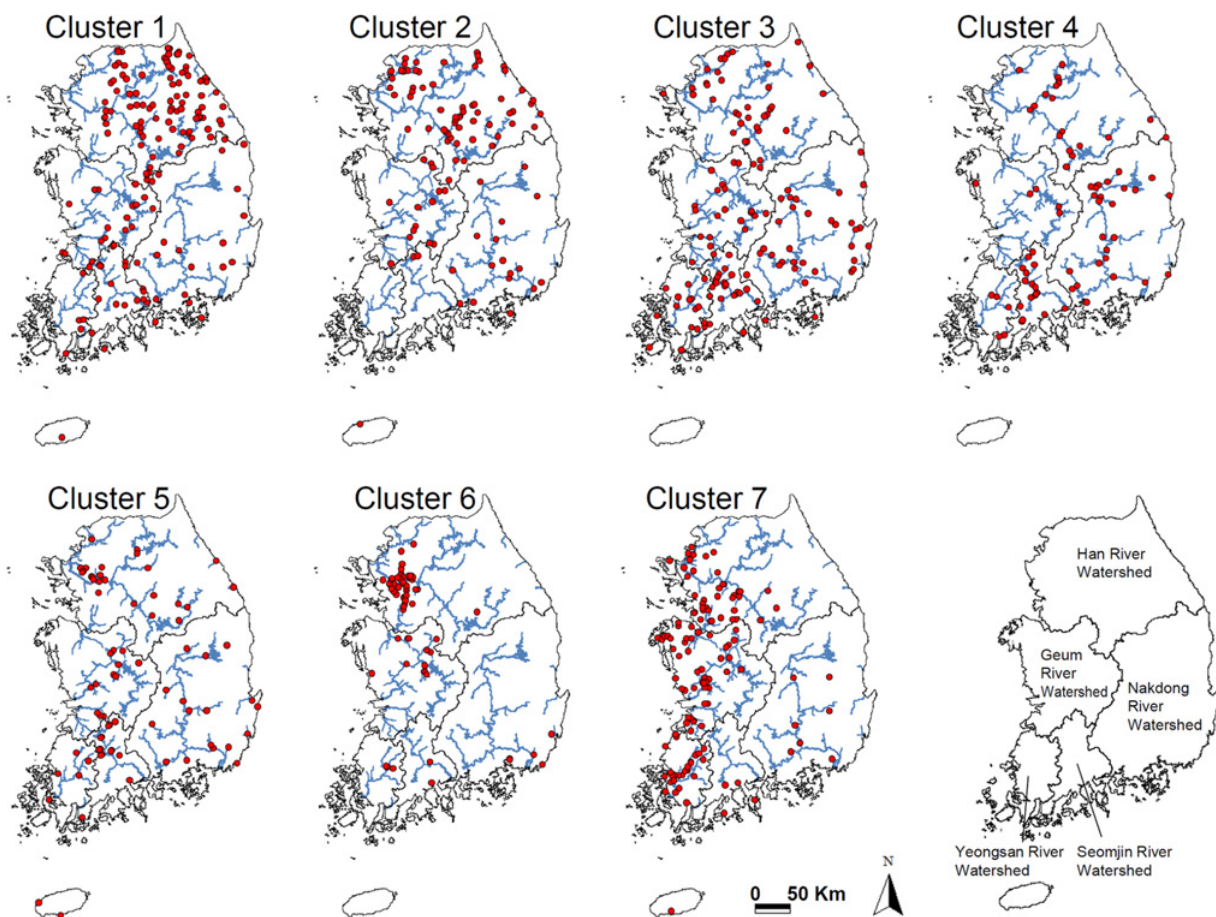

Fig. 4. Geological distribution patterns of seven different clusters defined in the SOM map trained with 27 environmental variables.
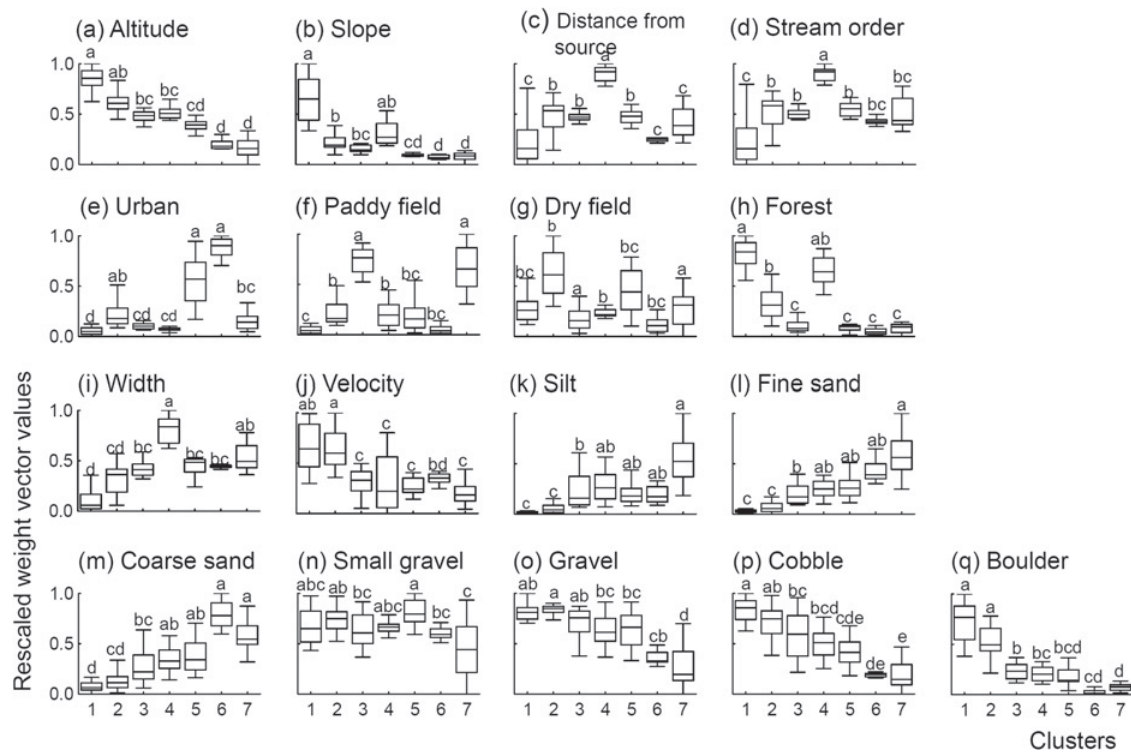
Fig. 5. Differences of environmental variables at different clusters defined in the SOM. Values were obtained from the weights of the trained SOM. (a–d) Sub-basin variables, (e–h) land cover types, (i, j) hydrological variables, and (k–q) substrate variables. Different alphabets indicate significant differences between the clusters based on Dunn's multiple comparison test ($P < 0.05$). (Boxplot: – median, □ 25–75%, I non-outlier range.)

variables. Distance from the source and stream order was low in high-altitude areas in cluster 1, and were high in low-altitude areas in cluster 7. Proportions of urban and paddy field were also different at different altitudes, displaying low values in high-altitude areas. Clusters 5 and 6 were characterized by a high proportion of urban, whereas clusters 3 and 7 contained a high proportion of paddy field. Clusters 2 and 5 showed high values of dry field, and cluster 4 displayed high values of distance from the source, stream order, and proportion of forest. Forest was high in cluster 1, while low in cluster 7. Current velocity was higher in high-altitude mountainous areas.

The ratio of small substrates, such as silt and sand, increased with decreased altitude, while large substrates, including cobbles and boulders decreased (Figs. 1 and 5). In addition, water quality tended to be worse with the decrease in altitude. For example, clusters 1 and 2, with high altitude and proportion of forest, showed relatively low values of water quality variables, such as conductivity, turbidity, and BOD, while clusters 5–7, with low altitude and high ratio of paddy field and urban, showed relatively high values of water quality variables (Fig. 6).

Species richness of diatoms in cluster 1, characterized by high altitude, was significantly lower than that in other clusters (Dunn's multiple comparison test, $P < 0.05$), whereas species richness of macroinvertebrates was the highest in cluster 1 and the lowest in clusters 6 and 7

(Fig. 7). Species richness of fish was the highest in cluster 4 and the lowest in cluster 6, displaying relatively low values in high-altitude areas. Other biological indices also displayed similar patterns with species richness of each assemblage, presenting different disturbance levels in different clusters. In all three assemblages, cluster 1 displayed less disturbed environmental conditions, whereas clusters 6 and 7 showed heavy disturbance. For diatoms, TDI was the lowest in cluster 1 and the highest in clusters 6 and 7 (Dunn's multiple comparison test, $P < 0.05$), whereas DAIpo displayed an inverse pattern of TDI due to their properties. ASPT and KSI in macroinvertebrates and K-IBIF and %_Sensi in fish also displayed similar patterns with those of diatoms.

Evaluation of important environmental variables

Clusters defined in the SOM with 27 environmental variables were well discriminated, showing high correct prediction (94.6 ($\pm$1.8%), mean ($\pm$S.E.)) through the random forest (Table 4). Based on MDG, land cover types were the most important variables in all clusters, except for cluster 7. Urban showed importance in clusters 5 and 6, paddy field in clusters 3 and 7, dry field in cluster 2, and forest in clusters 1 and 4, which were more dependent on sub-basin and hydrological variables than the other clusters. Meanwhile, water quality variables were more
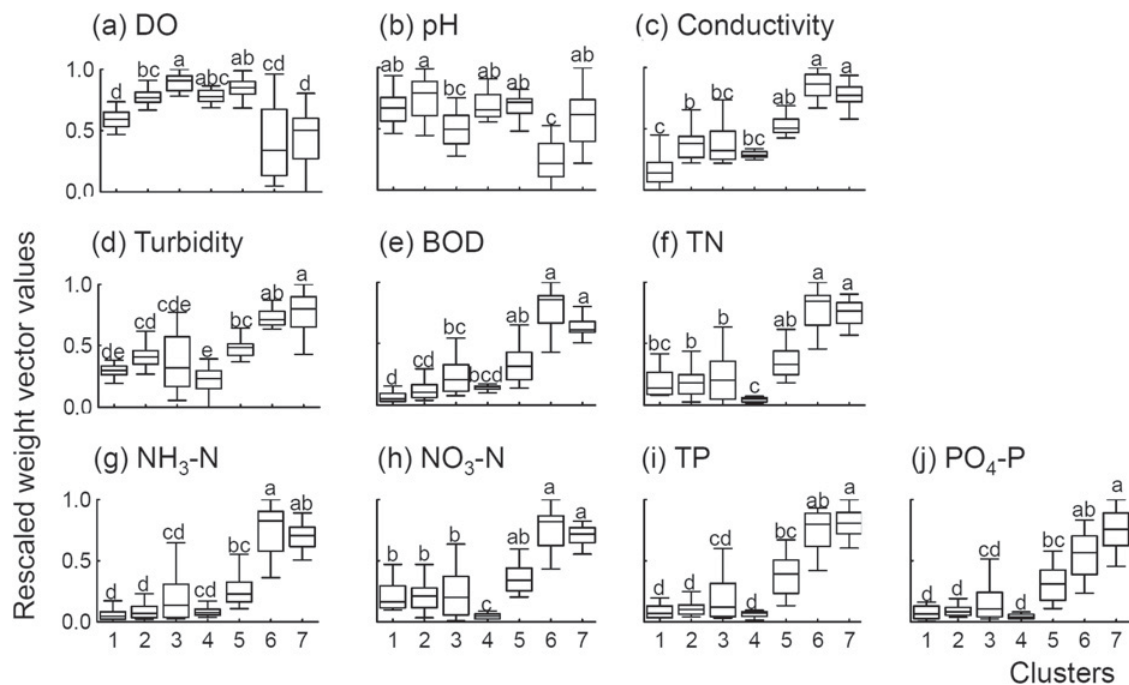
Fig. 6. Differences of estimated values of water quality variables at different clusters defined in the SOM. Values were obtained from the weights of the trained SOM. Different alphabets indicate the significant differences between the clusters based on Dunn's multiple comparison test ($P < 0.05$). (Boxplot: – median, □ 25–75%, Ɪ non-outlier range.)
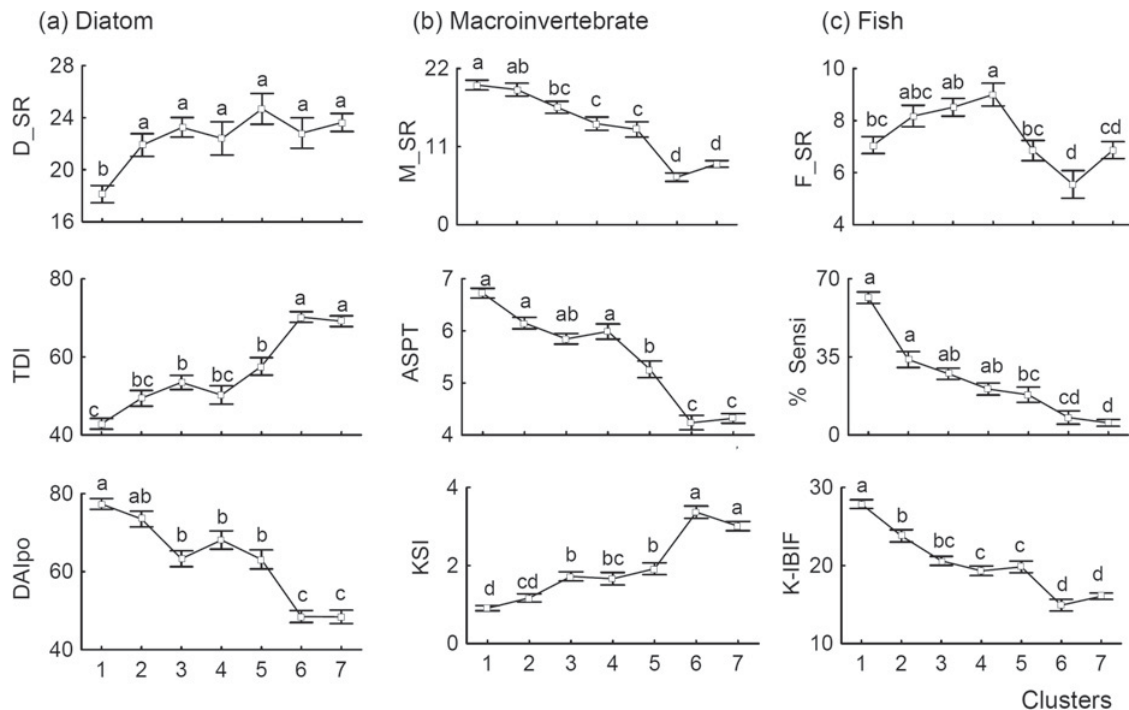

Fig. 7. Differences of biological indices in three different assemblages at different clusters defined in the SOM. (a) Diatom, (b) macroinvertebrate, and (c) fish indices. Different alphabets indicate the significant differences between the clusters based on Dunn's multiple comparison test ($P < 0.05$). Error bars represent standard errors. Abbreviations of the variables are shown in Table 2.

important than geo-hydrological variables for determination of clusters 5 to 7. Cobble in substrates was influential in clusters 2, 3, and 7, whereas gravel was influential in cluster 7. The importance of each variable was similarly reflected in the SOM results.

Prediction of biological indices in three assemblages

Biological indices could be predicted with environmental variables through the learning process of random forest, and the importance of environmental

Table 4. Prediction and variable importance of each cluster (top 10 variables represented in boldface type) using the random forest.

Environmental variables	Clusters						
	1	2	3	4	5	6	7
Sub-basin							
Altitude	**22.7**	**27.3**	5.8	**21.9**	11.6	4.4	22.0
Slope	**42.4**	**29.1**	4.0	14.5	11.8	2.8	6.8
Distance from the source	**14.0**	23.7	3.9	**92.8**	11.8	6.4	8.7
Stream order	**9.0**	8.0	1.3	**32.4**	3.7	1.4	3.6
Land cover							
Urban	**13.9**	**30.1**	**11.1**	17.7	**100.0**	**100.0**	**32.7**
Paddy field	**27.8**	**47.3**	100.0	28.7	**41.2**	27.9	**89.9**
Dry field	4.9	**100.0**	**9.2**	9.7	**24.6**	4.4	7.2
Forest	**100.0**	**87.9**	30.4	**100.0**	38.0	6.1	10.3
Hydrology							
Stream width	**18.3**	19.3	4.0	**67.4**	12.1	3.4	8.2
Current velocity	**8.7**	**48.1**	5.2	**57.5**	15.9	3.7	11.4
Substrate							
Silt	0.0	0.0	0.9	4.9	1.5	0.9	14.0
Fine sand	1.3	6.0	3.1	6.6	4.3	2.3	16.7
Coarse sand	1.4	8.5	2.7	8.2	8.2	7.1	8.0
Small gravel	1.7	8.9	1.8	6.3	6.6	1.6	6.6
Gravel	2.4	18.5	4.9	7.4	11.7	1.7	21.2
Cobble	4.1	**27.0**	6.2	14.5	6.1	2.1	**23.5**
Boulder	0.8	5.2	0.0	0.0	0.0	0.0	0.0
Water quality							
DO	3.8	17.9	**7.1**	14.8	**19.1**	5.2	16.4
pH	4.0	16.2	4.8	15.5	16.1	**7.7**	9.1
Conductivity	**27.0**	**34.9**	**8.6**	13.5	14.4	**9.9**	25.9
Turbidity	4.2	17.4	5.6	**24.2**	15.3	4.8	14.4
BOD	7.4	19.0	6.6	12.0	**24.9**	**29.3**	37.9
TN	3.4	16.5	**12.5**	42.9	**36.2**	**27.0**	**45.6**
NH$_3$-N	3.2	8.6	4.7	7.7	**23.6**	**27.3**	**40.0**
NO$_3$-N	3.5	14.9	**9.2**	43.2	29.2	22.3	29.2
TP	5.2	**25.4**	**18.9**	15.8	**22.0**	15.2	**100.0**
PO$_4$-P	2.7	17.9	**11.9**	11.4	18.9	10.2	**76.1**
Correct prediction (%)	95.0	95.3	95.6	94.2	94.6	97.4	93.9

variables was well evaluated; however, the prediction power was relatively low (R^2: 0.3–0.6) (Table 5). Overall, the contribution of altitude was high in the determination of all biological indices in the three different assemblages. Compared with other indices, land cover, except for forest and substrates, except for cobble, were less influential in the determination of biological indices. Meanwhile, the importance of environmental variables differed according to the different assemblages. For example, mainly water quality variables, such as DO, conductivity, BOD, and NH3-N were important for the determination of diatom indices. Macroinvertebrates were also affected by water quality variables, while water velocity was most influential to species richness in macroinvertebrates. In contrast to other assemblage indices, biological indices in fish were less influenced by water quality variables and the contribution of sub-basin variables was relatively high, compared with other assemblage indices.

Discussion

Streams are complex aquatic ecosystems in which a large number of environmental factors have been changed according to various spatial and temporal scales. These environmental factors show very diverse ranges, from physical, chemical, and biological characteristics in streams or rivers, to climate, land use, and geomorphology in watershed (Stevenson and Pan, 1999). Frissell *et al.* (1986) proposed the idea of a hierarchical, landscape-scale view of stream habitat, which provides a useful conceptual context for many aspects of both basic and applied stream ecology (Poff and Ward, 1990; Townsend, 1996; Poff, 1997). Poff (1997) developed a concept of multi-scale habitat filters and functional organization in streams, specifying a set of four habitat levels, including watershed, reach, channel unit, and microhabitat. The concept views environmental conditions as constituting filters through which species in the regional species

Table 5. Prediction of biological indices and evaluation of variable importance (top 10 variables represented in boldface type) using the random forest. Abbreviations of biological indices are shown in Table 2.

Environmental variables	Diatom			Macroinvertebrate			Fish		
	D_SR	TDI	DAIpo	M_SR	ASPT	KSI	F_SR	%_Sensi	K-IBIF
Sub-basin									
Altitude	**75.8**	**35.4**	**47.3**	**99.3**	**80.2**	**70.3**	35.1	**100.0**	**100.0**
Slope	20.9	7.6	10.5	11.1	8.5	8.4	13.7	**11.9**	**16.4**
Distance from the source	34.7	10.6	19.7	17.0	15.3	16.4	**100.0**	20.7	22.4
Stream order	12.6	6.3	9.2	5.1	5.0	6.2	**31.9**	8.1	8.7
Land cover									
Urban	28.9	9.1	11.6	20.0	12.4	15.1	**25.4**	7.0	9.2
Paddy field	**36.1**	9.1	13.8	16.1	7.7	10.0	**25.4**	6.7	7.4
Dry field	21.3	9.6	15.3	16.4	6.4	9.9	16.0	4.0	7.0
Forest	28.3	**20.4**	14.3	12.5	**17.6**	15.9	19.0	**22.1**	**26.4**
Hydrology									
Stream width	**44.5**	8.9	16.5	15.7	8.8	14.6	**66.9**	**40.0**	16.2
Current velocity	**52.8**	13.1	**23.2**	**100.0**	**50.1**	**47.8**	50.7	7.1	**36.1**
Substrate									
Silt	6.2	0.0	0.0	5.5	1.2	4.2	4.7	0.0	0.0
Fine sand	9.0	2.2	2.6	11.3	3.2	7.3	6.3	3.4	8.4
Coarse sand	16.5	7.2	10.2	12.3	4.2	12.8	9.3	4.0	5.4
Small gravel	18.2	5.1	9.4	11.3	4.1	6.7	8.1	3.2	3.6
Gravel	20.7	6.8	10.4	**43.0**	8.8	7.1	13.9	3.4	5.8
Cobble	15.8	10.8	11.3	**26.9**	11.0	**18.7**	22.3	**17.0**	22.2
Boulder	0.0	0.1	2.7	0.0	0.0	0.0	0.0	1.8	0.6
Water quality									
DO	**100.0**	13.0	**22.4**	**36.8**	18.9	18.0	**27.9**	9.3	10.4
pH	**54.5**	16.5	22.8	21.1	15.4	**24.9**	23.8	7.8	11.6
Conductivity	**72.9**	**100.0**	**100.0**	**79.3**	**76.7**	46.4	**52.7**	**56.1**	**93.7**
Turbidity	**45.3**	**21.4**	**32.5**	15.3	9.4	12.9	22.4	7.0	13.0
BOD	**72.3**	18.4	**35.3**	16.5	**20.9**	28.4	22.5	**15.3**	**37.3**
TN	34.6	13.9	21.7	**34.8**	**20.9**	20.9	25.3	7.4	**16.3**
NH3-N	**47.7**	17.1	**68.4**	8.6	7.0	12.6	17.6	**9.6**	12.4
NO3-N	32.6	**18.8**	18.5	**31.8**	16.3	24.8	32.6	6.5	11.0
TP	22.6	**61.2**	**96.8**	57.5	**100.0**	**100.0**	20.0	**21.6**	**33.0**
PO4-P	32.6	**32.2**	**44.3**	35.7	**56.3**	**62.7**	13.3	6.7	12.1
Prediction power (R^2)	0.3	0.4	0.4	0.5	0.6	0.4	0.3	0.5	0.5

pool, including all species in a region, must pass to be potentially present at a given locale (Tonn *et al.*, 1990). In this sense, various biological indices have been applied for presentation of the effects of integrated and hierarchical environmental factors, because biological indices provide information on the assessments of various environmental effects.

Therefore, in this study, we examined the relationships between the indices of three different biological assemblages (diatom, macroinvertebrate, and fish) and their environmental factors on multi-spatial scales. Our results paralleled the ideas of Frissell *et al.* (1986) and Poff (1997), showing strong effects of environmental variables on a larger scale to small-scale environmental variables.

Considering the concordance of different taxa, species richness of fish and macroinvertebrates displayed significant correlation, and diatoms showed a significant correlation with fish. However, diatoms did not show significant correlation with macroinvertebrates. Meanwhile, previous studies have revealed low assemblage

concordance in streams at within-basin scales (Allen *et al.*, 1999; Paavola *et al.*, 2003). These differences may be due to the differences of spatial scale in the study area. In other words, the data in our study came from a large spatial scale composed of five large river watershed and many small basins on the national scale. Thus, the current data support the idea that concordance could be found on a large spatial scale; however, concordance was not observed on a small scale, as proposed by Paavola *et al.* (2003). Allen *et al.* (1999) hypothesized a strong concordance with a similar body size of taxa. However, because fish and diatoms displayed concordance, while macroinvertebrates and diatoms did not show concordance, our study does not support this hypothesis. These patterns may be due to differences of ecological properties in each assemblage (Heino *et al.*, 2005). Therefore, our study supports the idea that concordance strength could be influenced by a combination of different factors, such as trophic relations or differences in biological traits between considered assemblages, and similar responses to

environmental gradient or position along a longitudinal gradient (Grenouillet *et al.*, 2007).

Species richness of each assemblage showed significant correlation with longitudinal gradient, although different responses were evident to species richness in each assemblage with longitudinal gradient (Vannote *et al.*, 1980). For example, altitude and slope showed significant correlation with all biological variables of the three assemblages. However, in diatom, species richness and altitude showed negative correlation. Although longitudinal pattern in the diversity of macroinvertebrate is still controversial, the greatest biotic richness will be found in medium-sized streams, such as 3rd–4th stream orders, showing high environmental variability (Melo and Froehlich, 2001). However, in our study, species richness of macroinvertebrates showed positive correlation with altitude, while it did not show correlation with distance from the source and stream order. Meanwhile, species richness of fish and diatoms were significantly correlated with distance from the source and stream order, indicating high species richness in large streams. This is because availability of solar energy for both photosynthesis and nutrients is essential to the distribution of diatoms (Allan and Castillo, 2007). In other words, the distribution of diatoms is limited in the headwater area in a forest with a closed canopy, while they are abundant in downstream regions with sufficient light and nutrients. In addition, the distribution of fish species along a longitudinal gradient has been well documented. Fish are more abundant in larger rivers than in smaller ones. This typical pattern has been explained in terms of habitat diversity (Gorman and Karr, 1978; Paavola *et al.*, 2006), differences of mobility and extinction, stream size (Paavola *et al.*, 2006), or sampling phenomena (Angermeier and Schlosser, 1989). It might also be influenced by available food resources, which are limited in the upper regions of streams.

In addition, interassemblage response to disturbances can vary as a result of differences in biological traits, including trophic structure, mobility, and longevity (Allen *et al.*, 1999). Therefore, diatoms have a strong correlation with local scale environmental factors, in particular, water quality variables (Hering *et al.*, 2006). Due to differences of trophic level from diatoms to macroinvertebrates to fish, their dependence on inorganic nutrients may decrease (Justus *et al.*, 2010). Therefore, diatom indices displayed much stronger responses, mostly to water quality variables than other variables. Therefore, nutrient availability may be one of the most significant factors to influence diatoms, which are the primary producers. Macroinvertebrates are more likely to respond to local habitat conditions, such as reach land cover, substrates, sedimentation, velocity, and pollution. This is because changes from forests to paddy fields or urban in land cover increase the sediment input to streams, which alters substrate composition (Lenat and Crawford, 1994; Quinn *et al.*, 1997). Meanwhile, fish, which have greater mobility and longevity, are more influenced by larger spatial conditions, such as catchment land cover, geology, and hydrology (Richards *et al.*, 1997;

Sponseller *et al.*, 2001; Black *et al.*, 2004; Flinders *et al.*, 2008).

Although all variables in multiple scales influence the prediction of aquatic assemblages, their relative contributions are different (Carter *et al.*, 1996; Townsend, 2003). Assemblage-specific responses to different types, levels, or scales of disturbance can result in low intercorrelation between diatoms, macroinvertebrates, and fish metrics (Flinders *et al.*, 2008). Therefore, there is growing recognition of the importance of larger-scale factors in the determination of the structure of aquatic communities (Poff and Allan, 1995; Townsend, 2003), supporting the notion that environmental factors operating at multiple scales affect aquatic communities (Black *et al.*, 2004). Local scale environmental conditions, such as nutrient inputs, sediment delivery, hydrology, and channel characteristics are influenced by large-scale conditions (Allan *et al.*, 1997). Regarding the concept of multi-scale habitat filters and functional organization in streams, Poff (1997) suggested that large-scale variables initially affect biota and that small-scale variables modify the influence on the remaining taxa. Large-scale variables may account for more of the variation in community composition than small-scale variables, and extreme factors at any scale can have an effect on community composition (Black *et al.*, 2004). Results of our study revealed different responses of fish, macroinvertebrates, and diatoms to variables at multiple scales, supporting the concept of multi-scale habitat filters (Poff, 1997). This may be due to differences of physical and chemical tolerances, as well as life history and biogeography among the three different assemblages (Townsend and Hildrew, 1994). In this sense, the use of multiple biological indices with more than one assemblage for the assessment of the biotic integrity of aquatic ecosystems would be warranted. Flinders *et al.* (2008) suggested the use of multi-metric indices for bioassessment of a stream's condition because they provide a broad measure of their aggregate impact and better support for assessment, and provide information on stressors not evaluated, as well in individual metric analyses. However, policy issues arise in defining a single impairment status from multiple assessments. Therefore, further studies are required for the development of methods of integrating information from multiple taxa groups and regional-specific calibration of integrated indices; although the integration of measures from multiple groups into a single assessment tool can be complicated both technically and from a policy point of view (Flinders *et al.*, 2008). In addition, the examination of taxa optima for specific environmental variables is needed to provide an approach that is more robust than traditional multi-metric approaches for the evaluation of both natural and anthropogenic effects on stream biota across multiple scales (Black *et al.*, 2004).

References

Allan J.D. and Castillo M.M., 2007. Stream Ecology: Structure and Function of Running Waters, 2nd edn., Kluwer Academic Publishers, Boston, 436 p.

Allan J.D., Erickson D.L. and Fay J., 1997. The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biol.*, *37*, 149–161.

Allen A.P., Whittier T.R., Larsen D.P., Kaufmann P.R., O'Connor R.J., Hughes R.M., Stemberger R.S., Dixit S.S., Brinkhurst R.O., Herlihy A.T. and Paulsen S.G., 1999. Concordance of taxonomic composition patterns across multiple lake assemblages: effects of scale, body size, and land use. *Can. J. Fish. Aquat. Sci.*, *56*, 2029–2040.

American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF), 2005. Standard Methods for the Examination of Water and Wastewater: Contennial Edition (Standard Methods for the Examination of Water and Wastewater), 21th edn., American Public Health Association, Washington, DC.

An K.G. and Lee E.H., 2006. Ecological health assessments of Yoogu stream using a fish community metric model. *Korean J. Limnol.*, *39*, 310–319.

Angermeier P.L. and Schlosser I.J., 1989. Species–area relationship for stream fishes. *Ecology*, *70*, 1450–1462.

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., Gerritsen J., Snyder B.D. and Stribling J.B., 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Algal, Benthic Macroinvertebrates, and Fish, 2nd edn., EPA 841-B-99-002, U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Beyer J., 1996. Fish biomarkers in marine pollution monitoring; evaluation and validation in laboratory and field studies, Academic Thesis, University of Bergen, Norway.

Black R.W., Munn M.D. and Plotnikoff R.W., 2004. Using macroinvertebrates to identify biota- and cover optima at multiple scales in the Pacific Northwest, USA. *J. N. Am. Benthol. Soc.*, *23*, 340–362.

Breiman L., 2001. Random forests. *Mach. Learn.*, *45*, 5–32.

Carter J.L., Fend S.V. and Kennelly S.S., 1996. The relationships among three habitat scales and stream benthic invertebrate community structure. *Freshwater Biol.*, *35*, 109–124.

DIN 38410, 1990. German standard methods for the examination of water, waste water and sludge: Biological-ecological examination of water (group M): Procedure for the determination of the saprobic index on the basis of benthic communities (M2), Deutsches Institut für Normung E.V., Berlin, 10 p.

Duong T.T., Feirtet-Mazel A., Coste M., Dang D.K. and Boudou A., 2007. Dynamics of diatom colonization process in some rivers influenced by urban pollution. *Ecol. Indic.*, *7*, 839–851.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy. *Off. J. Eur. Comm.*, *L327*, 1–72.

Flinders C.A., Horwitz R.J. and Belton T., 2008. Relationship of fish and macroinvertebrate communities in the mid-Atlantic uplands: Implications for integrated assessments. *Ecol. Indic.*, *8*, 588–598.

Frissell C.A., Liss W.J., Warren C.E. and Hurley M.D., 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environ. Manage.*, *10*, 199–214.

Gorman O.T. and Karr J.R., 1978. Habitat structure and stream fish communities. *Ecology*, *59*, 507–515.

Gregory S.V., Swanson F.J. and Mckee W.A., 1991. An ecosystem perspective of riparian zones. *BioScience*, *41*, 540–551.

Grenouillet G., Broe S., Tudesque L., Lek S., Baraillé Y. and Loot G., 2007. Concordance among stream assemblages and spatial autocorrelation along a fragmented gradient. *Divers. Distrib.*, *14*, 592–603.

Griffith M.B., Hillb B.H., McCormick F.H., Kaufmannd P.R., Herlihye A.T. and Sellef A.R., 2005. Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. *Ecol. Indic.*, *5*, 117–136.

Heino J., Paavola R., Virtanen R. and Muotka T., 2005. Searching for biodiversity indicators in running waters: do bryophytes, macroinvertebrates, and fish show congruent diversity patterns? *Biodivers. Conserv.*, *14*, 415–428.

Hering D., Johnson R.K., Kramm S., Schmutz S., Szoszkiewicz K. and Verdonschot P.F.M., 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. *Freshwater Biol.*, *51*, 1757–1785.

Infante D.M., Allan J.D., Linke S. and Norris R.H., 2009. Relationship of fish and macroinvertebrate assemblages to environmental factors: implications for community concordance. *Hydrobiologia*, *623*, 87–103.

Jain A.K. and Dubes R.C., 1988. Algorithms for Clustering Data, Prentice Hall, Englewood Cliffs, NJ, 304 p.

Johnson R.K., Furse M.T., Hering D. and Sandin L., 2007. Ecological relationships between stream communities and spatial scale: implications for designing catchment level monitoring programmes. *Freshwater Biol.*, *52*, 939–958.

Justus B.G., Petersen J.C., Femmer S.R., Davis J.V. and Wallace J.E., 2010. A comparison of algal, macroinvertebrate, and fish assemblage indices for assessing low-level nutrient enrichment in wadeable Ozark streams. *Ecol. Indic.*, *10*, 627–638.

Karr J.R., 1981. Assessment of biotic integrity using fish communities. *Fisheries*, *66*, 21–71.

Kelly M.G. and Whitton B.A., 1995. The trophic diatom index: a new index for monitoring eutrophication in rivers. *J. Appl. Phycol.*, *7*, 433–444.

Kohonen T., 2001. Self-Organizing Maps, 3rd edn., Springer, Berlin, 501 p.

Lenat D.R. and Crawford J.K., 1994. Effects of land use on water quality and aquatic biota of three North Carolina Piedmon streams. *Hydrobiologia*, *294*, 185–199.

Liaw A. and Wiener M., 2002. Classification and regression by randomForest. *R News*, *2*, 18–22.

Melo A.S. and Froehlich C.G., 2001. Macroinvertebrates in neotropical streams: richness patterns along a catchment and assemblage structure between 2 seasons. *J. N. Am. Benthol. Soc.*, *20*, 1–16.

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).

Paavola R., Muotka T., Virtanen R., Heino J. and Kreivi P., 2003. Are biological classifications of headwater streams concordant across multiple taxonomic groups? *Freshwater Biol.*, *48*, 1912–1923.

Paavola R., Muotka T., Virtanen R., Heino J., Jackson D. and Mäki-Petäys A., 2006. Spatial scale affects community concordance among fishes, benthic macroinvertebrates, and bryophytes in streams. *Ecol. Appl.*, *16*, 368–379.

Park Y.S., Chang J., Lek S., Cao W. and Brosse S., 2003. Conservation strategies for endemic fish species threatened by the Three Gorges Dam. *Conserv. Biol.*, *17*, 1748–1785.

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

Poff N.L.R., 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *J. N. Am. Benthol. Soc.*, *16*, 391–409.

Poff N.L.R. and Allan J.D., 1995. Functional organization of stream fish assemblages in relation to hydrological varia-bility. *Ecology*, *76*, 606–627.

Poff N.L.R. and Ward J.V., 1990. Physical habitat template of lotic systems: Recovery in the context of historical pattern of spatiotemporal heterogeneity. *Environ. Manage.*, *14*, 629–645.

Quinn J.M., Steele G.L., Hickey C.W. and Vickers M.L., 1997. Land use effects on habitat, water quality, periphyton, and benthic invertebrates in Waikato, New Zealand, hill-country streams. *N. Z. J. Mar. Freshwater Res.*, *28*, 391–397.

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

Rott E., 1991. Methodological aspects and perspectives in the use of periphyton for monitoring and protecting rivers. *In*: Whitton B.A., Rott E. and Friedrich G. (eds.), Use of Algae for Monitoring Rivers, Institut für Botanik, University of Innsbruck, Austria, 9–16.

Sponseller R.A., Benfield E.F. and Valett H.M., 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. *Freshwater Biol.*, *46*, 1409–1424.

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

Stevenson R.J. and Pan Y., 1999. Assessing ecological conditions in rivers and streams with diatoms. *In*: Stoermer E.F. and Smol J.P. (eds.), The Diatoms: Applications to the Environmental and Earth Sciences, Cambridge University Press, Cambridge, UK, 11–40.

Tang T., Cai Q. and Liu J., 2006. Using epilithic diatom communities to assess ecological condition of Xiangxi River system. *Environ. Monit. Assess.*, *112*, 347–361.

Tonn W.M., Magnuson M.R. and Toivonen J., 1990. Intercontinental comparison of small-lake fish assemblages: the balance between local and regional processes. *Am. Nat.*, *136*, 345–375.

Townsend C.R., 1996. Concepts in river ecology: pattern and process in the catchment hierarchy. *Arch. Hydrobiol. Suppl.*, *113*, 3–21.

Townsend C.R., 2003. Individual, population, community, and ecosystem consequences of a fish invader in New Zealand streams. *Conserv. Biol.*, *17*, 1, 38–47.

Townsend C.R. and Hildrew A.G., 1994. Species traits in relation to a habitat templet for river systems. *Freshwater Biol.*, *31*, 265–275.

Ultsch A., 1993. Self-organizing neural networks for visualiza-tion and classification. *In*: Opitz B., Lausen O. and Klar R. (eds.), Information and Classification, Springer-Verlag, Berlin, 307–313.

US EPA, 2002. Biological Assessments and Criteria Crucial Components of Water Quality Programs, U.S. Environmen-tal Protection Agency, Office of Water, EPA 822-F-02-006, Washington, DC.

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

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.

Watanabe T., Asai K. and Houki A., 1986. Numerical estimation of organic pollution of flowing water by using the epilithic diatom assemblage – Diatom Assemblage Index (DAIpo). *Sci. Total Environ.*, *55*, 209–218.

Won D.H., Jun Y.C., Kwon S.J., Hwang S.J., Ahn K.G. and Lee J.W., 2006. Development of Korean saprobic index using benthic macroinvertebrates and its application to biological stream environment assessment. *J. Korean Soc. Water Qual.*, *22*, 768–783 (in Korean with English summary).

Zelinka M. and Marvan P., 1961. Zur Präzisierung der biologischen Klassifikation der Reinheit fließender Gewässer. *Arch. Hydrobiologia*, *57*, 389–407.