Community patterns of benthic macroinvertebrates in streams in relation to temperature variation using the Self-Organizing Map

S.-H. Park¹, D.-H. Kim¹, W.-S. Cho¹, M. Bae², Y.-S. Park², Y.-E. Na³ & T.-S. Chon¹

¹Department of Biological Sciences, Pusan National University, Korea ²Department of Biology, Kyung Hee University, Korea ³Institute of Agricultural Science and Technology, Rural Development Administration, Korea

Abstract

Benthic macroinvertebrates were collected from 2,000 sample sites in the major river basins in South Korea from 1997 to 2002. In total 5 phyla, 9 class, 23 order, 111 family and 727 species of benthic macroinvertebrate were recorded during the survey period. Communities from the unpolluted sites (BMWP \geq 45) were selected and were trained by the Self-Organizing Map (SOM). The benthic macroinvertebrates were accordingly clustered based on topography of river basins in a larger scale, and subsequently according to temperature variation within the river basins in a smaller scale. The scope of species distribution was also illustrated through visualization of the SOM. Some selected species appeared to be locally adaptable while other species were widely distributed across the river basins. Patterning communities based on the SOM was efficient in monitoring species distribution and adaptability of the species in response to temperature variation.

Keywords: benthic macroinvertebrate, community, stream, temperature variation, self-organizing map.

1 Introduction

A trend in climate change is currently undeniable and ubiquitous [1], and the effects on global flora and fauna are significant in a wide range [2] resulting in



potential responses such as species extinct, *in situ* adaptation, and migration to tolerable areas [3]. There are also examples of species that may have succumbed to the pressures of contemporary climate change [4]. In correspondence with climate change, large-scale distributional changes in species have been recorded in a numerous account [e.g., 5].

Spatial distribution of benthic macroinvertebrates communities is closely associated with a variety of natural (e.g., temperature, pH) and anthropogenic (e.g., organic matter) factors [6]. Temperature is one of the key factors in determining distribution and abundance of benthic macroinvertebrates [7]. Although there are reports regarding longitudinal and seasonal variations [8], spatial distribution of benthic macroinvertebrates specifically related to the temperature gradient has not been studied extensively.

A difficulty for checking community response to environmental stress lies on complexity of data, since benthic macroinvertebrate consist of multi-variables (i.e., diverse taxa), varying in a non-linear fashion. The self-organizing map (SOM) is an efficient tool for mining non-linear data and has been extensively used for patterning communities since 1990s [9, 10]. The SOM efficiently classified benthic macroinvertebrate communities in polluted streams and illustrated community response to anthropogenic disturbances pertaining to the sample sites [7, 9]. In this study, we elaborated the SOM clustering and visualization in the large scale community data to address the distribution pattern of benthic macroinvertebrates in relation to topographic and temperature variation in the national scale.

2 Methodology

2.1 Ecological data

Benthic macroinvertebrates communities were sampled in relatively clean to intermediately polluted areas in South Korea. The samples were collected from the 2,000 sites in the major river basins in Korea (fig. 1) from 1997 to 2002 by using a Surber sampler (30 cm x 30 cm, $500 \mu \text{m}$).

Two longest rivers in South Korea, Nakdong and Han Rivers, were further divided into 2 and 3 regions, respectively, by considering topographic characteristics. The densities of collected individuals were converted to the unit area $(1m^2)$. The details for sampling benthic macroinvertebrates have been conducted according to the National Natural Environment Monitoring Protocol in Korea [11]. Community composition and other ecological information on the collected data could be referred to Park *et al.* [12].

Overall 5 phyla, 9 class, 23 order, 111 family and 727 species of benthic macroinvertebrates were recorded during the survey period. Shannon diversity index and the revised Biological Monitoring Working Party (BMWP) [13] were measured to present biological water quality. In order for the polluted communities not to be mixed with the unpolluted communities for analysis, the 584 sites with BMWP values equal to or over 45 were selected for training with the SOM. The rare species collected less than 10 sites during the survey period



were not included in the input data. Table 1 presents the number of sample sites, information on location, and biological indices for the major river basins used for analysis.



- Figure 1: The major river basins for collecting benthic macroinvertebrates in South Korea (H: Han River, N: Bukhan River (North trajectory in Han River), S: Namhan River (South trajectory in Han River), K: Geum River, J: Seomjin River, Y: Yeongsan River, U: North region in Nakdong River, D: South region in Nakdong River), M: Mountainous range.
- Table 1:The sample sites, location of the river basins, and diversity and
BMWP indices for benthic macroinvertebrate communities.

Rivers	Abbreviation	Sites (No.)	Latitude(°) (Range)	Longitude(°) (Range)	BMWP (Range)	Shannon Diversity (Mean ± S.D.*)
Bukhan	Ν	79	37.5-38.5	127.5-129.5	48.4-132.6	2.9±0.3
Namhan	S	81	37.5-36.8	127.5-129.5	47.9-121	3.1±0.3
Han	Н	49	36.8-38.5	126-127.5	51.3-150.8	2.8±0.3
Geum	К	69	35.8-36.8	126-128	47.2-138.6	2.9±0.3
Upstream of Nakdong	U	101	35.8-36.8	128-129.5	46.6-133.6	3.0±0.3
Downstream of Nakdong	D	42	34.5-35.8	128-129.5	58-147	2.9±0.3
Seomjin	S	136	34.5-35.8	127-128	46-134.2	2.8±0.3
Yeongsan	Y	33	34.5-35.8	126-127	46117	2.7±0.3

* Standard deviation

In order to check the longitudinal variation within the river systems in smaller scale, the data were rearranged after training the data for the whole river basins in South Korea. The data for south and north regions in the Nakdong River basin (138 species from 139 sites) were combined for training. Similarly the communities in the Namhan (S) and Han River basins (H) (167 species from 129 sites), and communities in the Bukhan (N) and Han River basins (H) (144 species from 127 sites) were separately combined for analysis. Beside the



Namhan and Bukhan River basins in mountainous area, only a few number of sample sites with high altitude were surveyed: four sites (over 600m) in the Nakdong River basin and one site (over 460m) in the Han River basin. The samples with high altitude in the Nakdong and Han River basins were not included in the data for analysis. For training the data within the same river basin, the species collected less than 5 sites were additionally excluded for analysis. In total 5 phyla, 9 class, 22 order, 107 family and 590 species were included in the input data for training.

2.2 Data analysis

We utilized the SOM for patterning macroinvertebrate communities [12, 14]. The output layer consists of $L \ge M$ computation nodes in the SOM. With each neuron being represented as j, the output layer is arranged in two dimensions for convenience of visual understanding. Suppose a community data containing S species, and the density of species, i, is expressed as a vector x_i . The vector, x_i is considered to be an input layer to the SOM. In the network each node, j, is supposed to be connected to each node, i, of the input layer. The connectivity is represented as the weights, w_{ij} (t), adaptively changing in each iteration of calculation, t. Initially, the weight is randomly assigned in small values. Each neuron of the network computes the summed distance between the weights and the distance d_j (t) at output node j network is calculated as shown below:

$$d_{j}(t) = \sum_{i=0}^{S-1} \left(x_{i} - w_{ij}(t) \right)^{2}$$
(1)

The input values with greatly different numerical values in densities are avoided for training. The data were transformed by natural logarithm in order 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.

The neuron responding maximally to a give input vector is 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 neurons are allowed to learn by changing the connecting weights in the manner to further reduce the distance between the weight and the input vector as shown below:

$$w_{ij}(t+1) = w_{ij}(t) + a(t)(x_i - w_{ij}(t))Z_j$$
(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. Detailed algorithm could be referred to Zurada [15], Chon *et al.* [16]. After training, the Ward's linkage method [17] was applied to the weights of the SOM for clustering the patterned nodes. Training and clustering were carried out under the MATLAB environment [18].

3 Results

3.1 Overall pattern

The sample sites were clustered accordingly after training by the SOM (figs. 2(a) and 2(b)). A majority of communities were separately clustered according to the river basins while other communities were observed together in different clusters. The samples from the river basins K, U, S and N were separately grouped in clusters 4, 8, 1 and 2, respectively. In addition, cluster 6 accommodated the samples from both south (D) and North (U) regions in the Nakdong River. The sample sites belonging to these clusters showed unique patterns with community compositions. The communities collected from J, H and Y, however, appeared together in a mixed manner in clusters 3, 5 and 7 (fig. 2(a)). Detailed comparison of communities in different clusters is beyond the scope of this study and will be reported elsewhere.



Figure 2: Community clustering in the 8 river basins by the SOM.
(a) Clusters of the sample sites on the SOM. The size of the symbols indicates the number of sample size grouped within the clusters.
(b) Dendrogram based on the Ward's linkage method.
(c) Latitude, longitude, altitude and temperature corresponding to the SOM shown in (a). The numbers in the vertical bar indicate the real values minus the minimum values (latitude: 34.5°, longitude: 126°, altitude: 0m, temperature: 10.8°).

The patterned nodes (fig. 2(a)) were further compared with latitude, longitude, altitude and temperature at the corresponding positions on the SOM (fig. 2(c)). Clustering of communities was not clearly matched to the profile of latitude in the large scale. This indicated that overall community compositions of benthic macroinvertebrates in South Korea are not in accordance with the southnorth longitudinal variation. In the profile of longitude, however, the gradient was observable: the values of longitude were higher in the left area, while the values were lower in the right area in the figure showing longitude (fig. 2(c)). This is understandable that topography in the Korean peninsula is divided into two regions in the east and west across the mountainous range longitudinally passing from north to south on the right side of the peninsula (fig. 1).



Consequently higher, colder mountainous areas are located in the east, while lower, warmer areas are placed in the west.

Altitude, however, did not clearly match community clustering (fig. 2(c)), while temperature was broadly in accordance with the profile of the longitude (fig. 2(c)). Temperature was generally higher in the area corresponding to the area of lower longitude (i.e., west area), while lower in the area matching to the area of higher longitude (i.e., east area) (fig. 2(c)). At the top and bottom left areas in fig. 2(c), values of temperature were especially lower, corresponding to the Bukhan (10.8°C±0.7) and Namhan (11.3°C±0.6) River basins, respectively.

3.2 Patterning within the river basins

Considering that the sample sites were clustered mainly according to the river basins in the large scale as shown in fig. 2(a). The two largest rivers in Korea, Nakdong and Han, were analyzed along the longitudinal axes. Fig. 3 shows clustering of the sample sites from the Nakdong River basin by the SOM. The communities were grouped along with the longitudinal axis (figs. 3(a), 3(b) and 3(d)). Cluster 1 mainly covered the sample sites from U (north region in the Nakdong River), while cluster 2 showed the middle range of the Nakdong River (fig. 3(c)), and cluster 3 corresponded to D (south region in the Nakdong River). The gradient of the latitude was associated with the vertical gradient in community clustering (figs. 3(a) and 3(d)). The profiles of longitude and altitude shown in fig. 3(d), however, did not match community clustering (fig. 3(a)). Temperature, in contrast, was in accordance with the gradient of the latitude and community clustering (fig. 3). The higher range in temperature was observed in the south region in the Nakdong area, while the lower range in temperature was shown in the north region in the Nakdong area (figs. 3(c) and 3(d)).



Figure 3: Community clustering in the Nakdong River basin by the SOM.
(a) Clusters of the sample sites on the SOM. (b) Dendrogram based on the Ward's linkage method. (c) The Nakdong River basin matching to clusters. (d) Latitude, longitude, altitude and temperature corresponding to the SOM shown in (a). The numbers in the vertical bar indicate the real values minus the minimum values (latitude: 34.51°, longitude: 128°, altitude: 0m, temperature: 12.6°).

Since two main trajectories are formed from the Han River (fig. 1), communities from the Bukhan and Namhan River basins were separately patterned. The main stream of Han River (H) was added to each trajectory. The clusters accordingly revealed the longitudinal differences in the Namhan and Han River basins (figs. 4(a) and 4(b)). The samples were divided into 3 clusters (fig. 4(a)). The cluster 1 presented the upstream region in the Namhan River (S), while cluster 2 matched the junction between the Namhan and Han Rivers. Cluster 3 showed communities from the Han River (H). Similar to the case of the Nakdong River basins (fig. 3(c)), communities were accordingly patterned along with the longitudinal axis of the river basin (figs. 4(a) and 4(c)). The gradient of the latitude was observed from upstream to downstream in the Namhan River basin (fig. 4(d)). Longitude also showed the gradient from left to right region in the Namhan River basin. This type of concurrent changes in both latitude and longitude are understandable by considering that the Namhan River runs diagonally toward south-east from the river mouth (fig. 1). Combining latitude and longitude together, diagonal changes in community patterns from south-east to north-west were demonstrated in the Namhan River basin (figs. 1, 4(c) and 4(d)). Consequently temperature matched to both latitude and longitude.



Figure 4: Community clustering in the Han and Namhan River basins by the SOM. (a) Clusters of the sample sites on the SOM. (b) Dendrogram based on the Ward's linkage method. (c) The Han and Namhan River basins matching to clusters. (d) Latitude, longitude, altitude and temperature corresponding to the SOM shown in (a). The numbers in the vertical bar indicate the real values minus the minimum values (latitude: 36.8°, longitude: 126°, altitude: 0m, temperature: 11.3°).

The communities in the Bukhan River basin were also divided into 3 clusters based on the SOM (figs. 5(a) and 5(b)). However, the community pattern was not formed in accordance with the longitudinal axis.

Cluster 2 presenting the upstream region in the Bukhan River (N) occupied the bottom position in community clustering, while cluster 1 corresponding to the junction of two rivers was placed in the top position. Cluster 3 matching the Han River basin (H) was located in the middle position in community clustering (fig. 5(a)). Latitude and longitude (fig. 5(d)) neither showed a clear matching to community clustering (fig. 5(a)) in this case. Altitude, however, was in accordance with community clustering. The high level in altitude (fig. 5(d)) corresponded to the top and bottom clusters shown in fig. 5(a). Profiles of temperature were in accordance with variation in altitude (fig. 5(d)). Since the Bukhan River is located in the mountainous area, and a numerous sample sites were selected in high altitude as stated above, the correspondence between altitude and community clustering was demonstrated in this river basin.



Figure 5: Community clustering in the Han and Bukhan River basins by the SOM. (a) Clusters of the sample sites on the SOM. (b) Dendrogram based on the Ward's linkage method. (c) The Han and Bukhan River basins matching to clusters. (d) Latitude, longitude, altitude and temperature corresponding to the SOM shown in (a). The numbers in the vertical bar indicate the real values minus the minimum values (latitude: 36.8°, longitude: 126°, altitude: 0m, temperature: 10.8°).

3.3 Distribution of species

The abundance profiles of the collected species were also accordingly displayed on the corresponding nodes in the SOM through visualization (figs 6–8). The species were broadly divided into broad and narrow groups in distribution. The species occurring in the narrow area were further divided into two types at the high and low temperature zones. Fig. 6(a) and 6(b) show the species distribution in narrow ranges in the Nakdong River basin for low and high temperature, respectively. Bottom region (fig. 6(a)) indicated lower latitude with high temperature, while upstream (fig. 6(b)) presented the opposite condition.

In the narrow range species such as *Ephemera strigata* and *Gammarus* sp were observed in the downstream zone (i.e., low temperature) (fig. 6(a)), while species such as *Hydropsyche* KUa,b and *Cincticostella castanea* were limitedly observed in the upstream zone (i.e., high temperature) (fig. 6(b)). Species including *Sieboldius albardae* were collected in a broad zone (fig. 6(c)).

Similar to the case of the Nakdong River basin, the scope of species distribution in the Namhan River basin showed the narrow (figs. 7(a) and 7(b)) and broad (fig. 7(c)) ranges. The collected species were divided into two narrow ranges matching the Han (bottom zone: fig. 7(a)) and Namhan (upper zone: fig. 7(b)) River basins. In the narrow range species such as *Laccotrephes*

japonensis and *Ischnura asiatica* were observed at the bottom zone (i.e., higher temperature) (fig. 7(a)). Species including *Zaitzevia nitida* and *Iron maculates* were narrowly presented in the upper zone in the Namhan River basin (fig. 7(b)). Species such as *Ephemera separigata* was collected in a broad range (fig. 7(c)).



Figure 6: Species distribution in the Nakdong River basin. Gray levels in vertical bars indicate density in no. of individuals per m².
(a) Species narrowly occurring in high temperature zone, (b) Species narrowly occurring in low temperature zone, and (c) Species occurring in a broad zone.



Figure 7: Species distribution in the Han and Namhan River basins. Gray levels in vertical bars indicate density in no. of individuals per m².
(a) Species narrowly occurring in high temperature zone, (b) Species narrowly occurring in low temperature zone and, (c) Species occurring in a broad zone.

The scope of species distribution in the Bukhan River basin also showed the narrow (figs. 8(a) and 8(b)) and broad (fig. 8(c)) ranges in distribution. The spatial distribution of the collected species in narrow ranges were divided into the Bukhan (bottom part: fig. 8(a)) and Han (middle part: fig. 8(b)) River basins. The species occurring in the Bukan River basin (fig. 8(a)) matched the area of high altitude in the SOM visualization (fig. 5(d)), while the species grouped in the Han River basin (fig. 8(b)) corresponded to low altitude in the SOM (fig. 5(d)). In the narrow range species such as *Ecdyonurus bajkovae* and *Drunella cryptomeria* were observed in the Bukhan River basin (fig. 8a), while species such as *Cheumatopsyche brevilineata* were collected in the Han River basin (fig. 8(b)). Species including *Ephemera orientalis* and *Ecdyonurus levis* were observed in a broad area (fig. 8(c)).



Figure 8: Species distribution in the Han and Bukhan River basins. Gray levels in vertical bars indicate density in no. of individuals per m².
(a) Species narrowly occurring in high altitude zone, (b) Species narrowly occurring high altitude zone, and (c) Species occurring in a broad zone.

4 Discussion and conclusion

In this study, we demonstrated that benthic macroinvertebrates communities trained by the SOM efficiently reflected effects of topographic difference and temperature variation (figs. 2–5). The spatial difference was accordingly illustrated by the SOM. Temperature variation was further identifiable with latitude (fig. 3), longitude and latitude (fig. 4), and altitude (fig. 5) in different river systems.

Through visualization the longitudinal differences in community compositions was demonstrated within the river basin systems. The narrow and broad ranges regarding temperature variation were efficiently elucidated in figs. 6, 7 and 8. The spatial ranges in relation to temperature variation may be useful for estimating the level of thermal tolerance. Considering that temperature may continuously rise according to climate change, species occurring at the narrow range in the warm areas may disperse upward. For instance, *Ephemera strigata* was limitedly observed in D would be the candidate species for upward distribution along with increase in temperature. In contrast the species presented at high latitude in low temperature, such as *Baetis ursinus* in N may extinct in South Korea along with the rise in temperature in the long period. Considering there are other numerous biological and abiological factors in determining position of target species, however, it is too early to conclude that these species would either flourish or extinct in South Korea. At least these results, however, indicate that the species stated above deserve attention for their abundance in relation to the long-term temperature increase.

In the Bukhan River basin, altitude was in accordance with the community patterns, while latitude and longitude did not clearly match community clustering (fig. 5). The Bukhan River basin is located near to the mountainous area as stated above. The effect of altitude difference on community patterns is consequently understandable. The results supported that temperature variation is not only dependent upon latitude and longitude but also dependent upon altitude. In the future the interrelationships among latitude, longitude and altitude could be studied in revealing temperature effects on communities in an integrative manner.

Seasonality may be another source of variation, especially in a temperature zone including Korea. In this study, however, the survey was fixed in the season covering late spring and summer. Seasonal differences could be further checked in the long term data. As is well known, pollution also plays an important role in determining community compositions [9]. The unpolluted sample sites were selected to reveal the natural effect of temperature variation by selecting the samples with the BMWP values over 45 in this study. In the future the effect of temperature on polluted communities needs to be further studied.

In conclusion, the SOM was efficient in patterning and visualizing community compositions and accordingly clustered benthic macroinvertebrates corresponding to the topography in a larger scale and to temperature variation in a smaller scale. The species distribution in relation to temperature variation was also identifiable with the SOM. Revealing the scope of spatial distribution would be important in characterizing community compositions and species adaptability to temperature variation. The SOM could serve as an efficient ecological map to illustrate community distribution in response to environmental stressors.

Acknowledgement

This study was carried out with the support of "Cooperative Research Program for Agricultural Science & Technology Development (project No.20080454000)", RDA, Republic of Korea.



References

- [1] Karl, T. R. and Trenberth K. E., Modern global climate change, *Science*, 302, pp. 1719-1723, 2003.
- [2] Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C. and Pounds, J. A., Fingerprints of global warming on wild animals and plants, *Nature*, 421, pp. 57-60, 2003.
- [3] Coope, G. R., *Insect faunas in ice age environments: why so little extinction? In: Extinction Rates* (eds Lawton, J. H., May, R. M.), pp. 55-74, Oxford University Press, New York, 1995.
- [4] Pounds, J. A., Bustamante, M. R., Coloma, L. A., Consuegra, J. A., Fogden, M. P. L., Foster, P. N., Marca, E. L., Masters, K. L., Merino-Viteri, A., Puschendorf, R., Ron, S. R., Sánchez, G. A., Still, C, J. and Young, B. E., Widespread amphibian extinctions from epidemic disease driven by global warming, *Nature*, 439, pp. 161-167, 2006.
- [5] Hickling, R., Roy, D. B., Hill, J. K., Fox, R. and Thomas, C. D., The distributions of a wide range of taxonomic groups are expanding polewards, *Global Change Biology*, 12, pp. 1-6, 2006.
- [6] Cao, Y., Bark, A. W. and Williams, W. P., Analysing benthic macroinvertebrate community changes along a pollution gradient: A framework for the development of biotic indices, *Water Research*, 31, pp. 884-892, 1997.
- [7] David, D. L., Jerry, L. W. and Daniel, G. P., The effect of trout farm effluent on the taxa richness of benthic, *Aquaculture*, 147, pp. 37-55, 1996.
- [8] Leunda, P. M., Oscoza j., Mirandaa, R. and Ariñoa, A. H., Longitudinal and seasonal variation of the benthic macroinvertebrate community and biotic indices in an undisturbed Pyrenean river. *Ecological Indicators*, 9, pp. 52-63, 2009.
- [9] Song, M. Y., Hwang H. J., Kwak, I. S., Ji C. W., Oh, Y. N., Youn, B. J. and T. S. Chon, Self-organization mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation, *Ecological modeling*, 203, pp. 18-22, 2007.
- [10] Park, Y. S., Verdonschot, P., Chon, T. S. and Lek, S., Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network, *Water Research*, 37, pp. 1749-1758, 2003.
- [11] Ministry of Environment, The national natural environment monitoring protocol, Ministry of Environment, Seoul, 1997.
- [12] Park, Y. S., Song, M. Y., Park, Y. C., Oh, K. H., Cho, E. and Chon, T. S., Community patterns of benthic macroinvertebrates collected on the national scale in Korea, *Ecological Modeling*, 203, pp. 26-33, 2007.
- [13] Walley, W. J. and Hawkes, H. A., A computer-based development of the Biological Monitoring Working Party score system incorporating abundance rating, biotope type and indicator value, *Water Research*, 31(2), pp. 201-210, 1997.
- [14] Kohonen, T., Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, 43, pp. 59-69, 1982.

- [15] Zurada, J., Introduction to artificial neural systems, ISBN, 1992.
- [16] Chon, T. S., Park, Y. S., Moon, K. H. and Cha, E. Y., Patterning communities by using an artificial neural network, *Ecological Modelling*, 90, pp. 69-78, 1996.
- [17] Ward, J. H., Hierarchical grouping to optimized an objective function, *Journal of the American Statistical Association*, 58, pp. 236–244, 1963.
- [18] The Mathworks Inc., Matlab version 6.1, Massachusetts, 2001.

