Development of a benthic macroinvertebrate predictive model based on the physical and chemical variables of rivers in the Republic of Korea

Jeong-Ki Min, Hyoju Lee and Dongsoo Kong

ABSTRACT
Predictive models for the benthic macroinvertebrate community based on environmental variables facilitate the identification of the organisms expected to inhabit an area according to the target environmental conditions when restoring rivers. In this investigation, a biotic community predictive model was developed using benthic macroinvertebrate and environmental variable data collected from 1,210 sites in the Republic of Korea from 2010 to 2020. The sites were classified into six groups according to Two Way Indicator Species Analysis (TWINSPAN) and based on their individual abundance/m² of benthic macroinvertebrates. The TWINSPAN groups were related to 14 variables by stepwise multi-discriminant analysis. The relative importance of the environmental variables that classified each TWINSPAN group was in the order of mean diameter of particle size, catchment area, altitude, velocity, total phosphorus, latitude, pH, longitude, conductivity, water depth, suspended solids, biochemical oxygen demand, stream order, and total nitrogen. Discriminant functions 1–4 showed statistically significant and a predictive model was developed using functions 1 and 2 based on Wilks’ lambda values. The fit of the derived model was confirmed using Sørensen similarity (number of taxa) and Bray–Curtis dissimilarity (individual abundance/m²) analyses between the predicted organisms and those observed at the sites. The distributions of similarity and dissimilarity that were confirmed by stream type ranged from 0.60 to 0.72 and 0.46–0.56, respectively, based on the mean. Based on the predicted and observed values, the ratio of shredders and scrapers to collectors showed similar results overall for each stream type. The predictive model derived using nationally managed available data is expected to be applicable to stream and river restorations in the future, as it provides a statistical assessment of the biotic communities that are expected to inhabit a given environment.

KEY HIGHLIGHTS POINTS
- A biotic community predictive model is presented. The model was developed using benthic macroinvertebrate and environmental variables.

CONTACT Dongsoo Kong, dskong@kgu.ac.kr
Department of Life Science, Kyonggi University, Suwon-si, 16227, Gyeonggi-do Province, Republic of Korea.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
variable data collected from 1,210 sites across the Republic of Korea from 2010 to 2020.

- The purpose of the model is to identify communities that should be present in river environments after restoration under modified environmental conditions.
- The model can function on a larger scale to address the increasing need for river restoration from a broader perspective.
- Model usage will provide successful and sustainable results and meet the needs of policy makers to restore riverine environments.

1. Introduction

Globally, river ecosystems are threatened by the influx of pollutants around watersheds, the changes in hydraulic and hydrological systems that occur for the prevention of natural disasters, and the human use of water resources (such as drinking water, irrigation, food resources such as fishes). (Boon 1988; Malmqvist and Rundle 2002; Haapala et al. 2003; Dudgeon et al. 2006). Especially, industrialization and urbanization have progressed rapidly to enable economic growth, but have led to an increase in development and land use in river basins, which has resulted in the deterioration of ecological integrity and reduced biodiversity in stream ecosystems (Boon 1988; Bazinet et al. 2010; Cuffney et al. 2010). These threats to river ecosystems can reduce biodiversity and lead to the extinction of certain aquatic organisms (Strayer and Dudgeon 2010). It is thus important to improve and maintain the ecological health of damaged rivers through restoration (Lake et al. 2007; Craig et al. 2008; Selego et al. 2012). For this reason, efforts have been made to establish various conservation strategies and management techniques to maintain sustainable river ecosystems and biodiversity, but there were difficulties due to the lack of knowledge and information regarding restoration ecology (Bucher et al. 1997; Dudgeon 2000; Strayer 2006). Thus, restoration ecology has recently emerged as an important scientific field for environmental managers and decision makers, and is of great public interest, leading to an increase in river restoration with an ecological focus (Ormerod 2003; Louhi et al. 2011). In other words, the maintenance of healthy river ecosystems and the restoration of partially damaged rivers has become a major focus of environpolitics, and consequently, a large amount of effort has been invested in river restoration (Brown 2000; Koehn et al. 2001; Schanze et al. 2004). For example, in the international context, the European-Union Water Framework Directive (WFD) recommends that member states restore rivers of poor quality (WFD. 2000; Carvalho et al. 2019) and the United States has aimed to restore and preserve the physical, chemical, and biological integrity of river ecosystems since the Clean Water Act was enacted in 1972 (Selvakumar et al. 2010). In addition, measuring the effectiveness of past projects is important to gain a better understanding of restoration practices and to help ensure that restoration plans will be successful (Selvakumar et al. 2010; Selego et al. 2012). Accordingly, there have been an increasing number of investigations assessing the specific guidelines for the formation of appropriate physical and chemical habitats in restorations (Palmer et al. 2005; Cramer 2012; ME. 2014; Yochum 2018), and whether ecological restorations have been successfully conducted with freshwater macroinvertebrates and fishes (Haapala et al. 2003; Selvakumar et al. 2010; Selego et al. 2012; Kim et al. 2016). However, worldwide, most river restorations have been small one-time projects for channel design, riparian zones, or floodplains...
(Bond and Lake 2003; Haapala et al. 2003; Selego et al. 2012; Clark and Montemarno, 2017), and extensive restoration of entire watershed is very rare (Lake et al. 2007).

The goal of river restoration is to improve the negative impacts that humans have had on an ecosystem and return the river to a state that remains natural without human intervention (Gørtz 1998; Selego et al. 2012). However, such processes may change existing aquatic communities (Berkman et al. 1986; Tikkonen et al. 1994; Merz and Ochikubo Chan 2005) and the effects on the aquatic organisms may differ depending on the purpose of the restoration (Lake et al. 2007). To understand what is required of restoration guidelines, we must first understand what we would classify as a successful restoration. In most cases, the results of river restoration were assessed based on changes in species composition and individual abundance or species diversity and richness before and after restoration (Miller et al. 2010; Merz and Ochikubo Chan 2005; Kim et al. 2016).

Consequently, a river restoration was classified as successful if the diversity of substrates had increased, various microhabitats were provided to benthic macroinvertebrates, and organisms with sensitivity to pollution inhabited the water and improved the quality. If the main purpose of restoration is to increase biodiversity by improving physical and chemical variables in the target reach, it will be possible to assess it by comparing the biotic communities before and after restoration, and if biodiversity has increased, the river restoration is classified as successful. However, it is difficult to judge the success of restorations performed for a specific taxon regarding the restoration of another taxon. For example, research on benthic macroinvertebrates in the restoration reach performed to increase the individual abundance of a specific fish will have no other method than just comparing changes in the community structures before and after restoration (Merz and Ochikubo, 2005).

River restoration can be performed by controlling the environmental conditions that affect biological interactions, and after restoration, it seeks to establish specific organisms that prefer the target environmental conditions and will be able to sustain their populations (Lake et al. 2007). This can be achieved by providing organisms with appropriate food sources and shelter from natural disturbances (Lancaster and Hildrew 1993). For example, Haapala et al. (2003) confirmed that by improving the retention capacity of a streambed in Rutajoki, a third-order stream located in central Finland, the abundance of benthic macroinvertebrates increased compared to other reaches, as organic matter such as fallen leaves accumulated in the reach. This was due to an overall increase in the number of shredder and detritivores feeding on the fallen leaves and several predators (e.g., Plecoptera) feeding on these organisms. After setting specific goals for ecosystem processes or organisms to be restored in this way, systematically planned river restorations can create ecological characteristics to meet these goals. However, other river restorations have been conducted with the aim of increasing biodiversity by increasing the heterogeneity of microhabitats overall, rather than with specific goals, and few cases are monitored after restoration (Selvakumar et al. 2010; Louhi et al. 2011; Kim et al. 2016).

Understanding how river organisms, environmental variables, and processes affect each other is key to understanding an aquatic ecosystem as a whole and have a successful restoration (Nelson and Lieberman 2002; Sandin 2003; Li et al. 2012). This means that to successfully perform river restoration, it is necessary to identify the environmental conditions preferred by the target organisms. Additionally, they can also be interpreted to estimate the organisms expected to inhabit the target environment when the river is restored. The habitat and the resulting abundance of benthic macroinvertebrates by taxa will depend on a wide range of geographical features and ecological processes, primarily on the availability of adequate habitats and food resources (Lake et al. 2007). However,
accurate information on habitat and food availability based on a wide range of river environments and corresponding ecosystem processes may be difficult to obtain (Clarke et al. 2003). In addition, based on the river continuum concept, the biotic communities change gradually rather than discretely (Vannote et al. 1980; Min et al. 2019). Thus, it is difficult to determine which organisms are expected to inhabit a given environment. This problem can be solved through biotic predictive models such as United Kingdom’s River Invertebrate Prediction and Classification System (RIVPACS; Clarke et al. 2011), Australia’s Australian River Assessment System (AUSRIVAS; Smith et al. 1999), and Spain’s Mediterranean Prediction And Classification System (MEDPACS; Poquet et al. 2009). These models have been applied to aquatic environmental assessments using benthic macroinvertebrate prediction. It is believed that this approach can answer, whether specific species are more important than others when restoring certain ecosystem processes (Lake et al. 2007).

RIVPACS is a predictive system developed for biological environmental impact assessment (Cox et al. 1997), initially aimed at developing biotic classifications of reference streams based on benthic macroinvertebrates and whether the benthic macroinvertebrate communities could be predicted based on the environmental characteristics of the sites (Wright et al. 1984). The goal of the statistical predictive model is to suggest standardized biotic communities based on measurable environmental variables in many rivers. According to Clarke et al. (2003), the environmental variables used in the predictive model only need to have high predictive power as they show a high correlation with biotic communities, and the variables do not necessarily have to be variables of community composition and causality. Therefore, procedures were required to select reference streams and environmental variables for prediction, as well as to use taxa (identification level) (Clarke et al. 2003; 2011). The reason the selection of reference streams is included in the RIVPACS development process is to assess the current status of the site by predicting the biotic communities expected to inhabit undisturbed streams through specific available environmental variables independent of natural and artificial disturbances (Clarke et al. 2011).

Our biotic community predictive model for river restoration was derived based on the concept of RIVPACS, the model construction entirely followed the method of Wright et al. (2000). Thus, predicting biotic communities using the environmental characteristics of the sites was derived using: 1) grouping of sites based on biotic attributes, 2) probability that the sites belong to each group classified biologically, and 3) frequency of occurrence and abundance (or density) of each taxon in each group (Wright et al. 2000). However, our biotic predictive model differed in part from that of RIVPACS, whose main purpose is to assess rivers, as its main purpose is to probabilistically suggest the biotic communities expected to inhabit a given environment after stream or river restoration. Therefore, a separate reference stream for the predictive model was not selected and the environmental variables with which the predictive model is used, were used not only for invariant variables and for all variables currently managed, established, and used by the National Institute of Environmental Research.

Benthic macroinvertebrates, a biotic variable used in RIVPACS, have been widely used as aquatic indicator organisms for the biotic assessment of rivers due to their high species diversity, long lifespan, bottom-dwelling, and sensitivity to habitat disturbance characteristics, and they have been successfully used for short-term and long-term river environment monitoring (Rosenberg et al. 1986; Resh and Jackson 1993; Karr 1999; Smith et al. 1999). Furthermore, because they are part of the river food web, the role of benthic
Macroinvertebrates is important to understand the various processes that occur in river ecosystems (Li et al. 2012). The Republic of Korea, as in other countries, has developed an increased awareness of the need for river ecosystem restoration and preservation. Research into river restoration has predominantly been conducted by National institutes (Cho et al. 2010; ME. 2014; Kim et al. 2016), and some rivers have been restored on a trial basis (Kim 2007; Kim et al. 2016). However, since the concept of river restoration was not clearly established at the time of these previous restorations, they focused more on the river landscape than on the river ecosystem (Woo 2009). Since these original restorations, a technical guideline for river restoration projects has been proposed, focusing first on ecological restoration by synthesizing overseas river restoration guidelines and considering the characteristics of the rivers in the Republic of Korea (ME. 2014). The guidelines recommend the selection of flagship species (such as birds, amphibians, fish, and benthic macroinvertebrates) that can represent local characteristics or represent the river’s environment after restoration. However, they only recommend the selection of flagship species and do not suggest which organisms are suitable for the environment after restoration or which environmental conditions are suitable for a particular species to inhabit. This can be determined by predicting which organisms are expected to inhabit the target environment conditions after restoration, but documented cases of this are hard to find. Therefore, the purpose of this study was to develop a predictive model for benthic macroinvertebrate communities at the genus level that are expected to inhabit the target environment when restoring rivers in the Republic of Koreas.

2. Materials and methods

2.1. Study area

This study involved 1,210 sampling sites (13,354 sampling units), and these included the mainstems, tributaries, and other independent streams of five major river watersheds (Han River, Nakdonggang River, Geumgang River, Yeongsangang River, and Seomjingang River) on a national scale in the Republic of Korea (Figure 1). The total area of the Republic of Korea is approximately 100,339 km², and it has a low-altitude mountainous topography that accounts for approximately 60% of the total land area (Kim et al. 2017). High-altitude mountains are located mainly in the eastern region, creating a steep slope, with a low-lying plain in the western area (highest east and West low) as the altitude decreases (Lee et al. 2011; Li et al. 2012). There are more mountain streams with high or low altitudes than lowland plain streams due to topographical features (Bae et al. 2003).

2.2. Data collection

This study used the results of the Nationwide Aquatic Ecological Monitoring Program (NAEEMP) conducted by the Ministry of Environment and the National Institute of Environmental Research from 2010 to 2020. Environmental variables and sampled benthic macroinvertebrates were measured according to the NIER (2017) and NIER. (2019) guidelines.

A total of 15 environmental variables could be used as they were collected and managed nationwide for the development of a predictive model (Table 1). These variables were measured before sampling the benthic macroinvertebrates at each sampling site. The order of the stream corresponding to the size of the stream was determined using a
Figure 1. Location of the 1,210 sampling sites (circles) in the Republic of Korea. Black thick lines indicate the four major rivers (Han, Nakdong, Geum, and Yeongsan-Seomjin) watersheds. Thin blue lines in each black line indicate the main, tributary, and other independent streams of each watershed.

Table 1. Four categories for each candidate environmental variable in the biotic predictive model.

| Categories               | Environmental variables                        | Units |
|--------------------------|------------------------------------------------|-------|
| Stream size              | Stream order (Order)                            |       |
|                          | Catchment area (Area)                           | km^2  |
|                          | Stream width (Width)                            | m     |
| Geographical feature     | Altitude                                       | m     |
|                          | Latitude                                        | –     |
|                          | Longitude                                       | –     |
| Microhabitat             | Velocity                                        | cm/s  |
|                          | Water depth (Depth)                             | cm    |
|                          | Mean diameter of particle size in substratum (Diameter) | \( \phi_m \) |
| Water quality            | pH                                              | –     |
|                          | Biological oxygen demand (BOD_5)                | mg/L  |
|                          | Suspended solid (SS)                            | mg/L  |
|                          | Total nitrogen (TN)                             | mg/L  |
|                          | Total phosphorus (TP)                           | mg/L  |
|                          | Conductivity                                    | \( \mu S/cm \) |
1:120,000-scale map according to the criterion of Strahler (1957). The catchment areas were calculated based on the Korean Reach File provided by NIER. (2015). Stream width was determined as the distance from bank to bank using a laser range finder (Bushnell Inc., Overland Park, KS, USA). The altitude, latitude, and longitude corresponding to the geographical features of the streams were measured using GPS equipment (Triton 500, Magellen Inc., San Dimas, CA, USA). Water depth, velocity, and composition of the streambed which are related to microhabitat were measured at the points at which the benthic macroinvertebrates would be sampled at each sampling site, and the mean values were used. The water depth was measured as the surface of the vertical distance from the water surface to the streambed. The velocity was determined by using a flow meter (Flow-Mate Model 2000 or 3000-LX, Swoffer Instruments, Ins., Tukwila, WA, USA) or by applying the Craig method (Craig 1987). The composition of the streambed was first classified into boulder ($D_m > 256$mm, $\phi < -8$; $D_m$ : diameter (mm), $\phi = -\log_{10}D_m$; Bunte and Abt 2001), cobble ($D_m = 64 \sim 256$mm, $\phi = -8 \sim -6$), pebble ($D_m = 16 \sim 64$mm, $\phi = -6 \sim -4$), gravel ($D_m = 2 \sim 16$mm, $\phi = -4 \sim -1$), and sand ($D_m \leq 2$mm, $\phi \geq -1$) according to the criterion of Cummins (1962), and the ratio of each type of substrate that occupied each sampling area was measured at the sampling points with the naked eye (NIER. 2019). The median diameter of each substrate type (boulder, $-9$; cobble, $-7$; pebble, $-5$; gravel, $-2.5$; sand, 0) and the weighted average of the ratio of the area occupied by each substrate at the sampling points were used to calculate the mean diameter of the particle size in the substratum. The variables corresponding to water quality, pH, and conductivity were measured using multiprobe portable meters (YSI 6920, YSI Inc., Yellow Springs, OH, USA or Horiba U-22XD, Kyoto, Japan) and biochemical oxygen demand (BOD$_5$), suspended solid (SS), total nitrogen (TN), and total phosphorus (TP) were analyzed according to the water pollution process test standards (NIER 2021).

The benthic macroinvertebrates were sampled in two main ways. A Surber sampler ($30 \times 30$ cm, 1 mm mesh) was used for quantitative sampling and was conducted mainly on the riffle at each sampling site, and as the stream size increased, such as non-wadable streams, at sites where could not sampling at riffle, accessible water depth area was targeted (mainly pool). The sampling of some non-wadable streams using Ponar grab ($20 \times 20$ cm) was conducted at the center of a river. Quantitative sampling was conducted annually at the sites by collecting three replicates (Surber sampler, $0.27$ m$^2$ or Ponar grab, $0.12$ m$^2$ for each sampling unit) in spring and autumn (some are sampled only once in spring or autumn due to inaccessibility, etc.). Three replicates for each sample were placed in one collection bottle and fixed in 95% ethanol, and then transferred to the laboratory, where samples were separated from detritus and inorganic matter and then stored in 80% ethanol after identification. All specimens were identified at the genus level if possible, but taxa with limited information (e.g., Oligochaeta, Hemiptera, Coleoptera, Diptera) were identified at the family level (Yoon 1995; Kwon et al. 2001). The individual abundance for each taxon sampled was counted and converted to an individual abundance/m$^2$ as the quantitative sampling area for each sampling tool was different.

### 2.3. Data analysis

Principal component analysis (PCA) was performed on 15 environmental variables to confirm the relationship between environmental variables before analysis with the predictive model. Because PCA requires that the used environmental variables used to follow the normal distribution, they were converted into $\ln(x+1)$ when the raw data value for each variable itself did not have a normal distribution. Then, based on the variable values
showing the normal distribution, the range of each environmental factor was adjusted as shown in Equation (1) based on the maximum and minimum to equalize the range of each environmental variable to 0–1.

$$\text{Adjusted environmental value} = \frac{V_x - V_{\text{min}}}{V_{\text{max}} - V_{\text{min}}}$$  \hspace{1cm} (1)

Overall, the predictive model was developed as suggested by Moss et al. (1987) and Wright et al. (2000). Therefore, a hierarchical cluster analysis was performed using Two-Way Indicator Species Analysis (TWINSPAN; Hill 1979) to group the 1,210 sites based on the similarity of their communities using the individual abundance/m² of benthic macroinvertebrates. Cluster analysis may have different results depending not only on the algorithms used for the analysis but also on the occurrence characteristics of the organisms. Therefore, as attempted in Min et al. (2019), the individual abundance/m² for all taxa was adjusted to Equation (2), reducing the influence of some specific dominant taxa and transforming the individual abundance/m² within the same range. Then, to confirm whether the difference in environmental variables for each TWINSPAN group was statistically significant, the Kruskal–Wallis test, a nonparametric statistic, was performed using the raw data of each environmental variable.

$$\text{Adjusted abundance value} = 0.5 + \frac{\ln n - \ln n_{\text{min}}}{\ln n_{\text{max}} - \ln n_{\text{min}}} = 0.5 + \log\left(\frac{n_{\text{max}}}{n_{\text{min}}}\right)\left(\frac{n}{n_{\text{min}}}\right)$$  \hspace{1cm} (2)

The TWINSPAN groups for the 1,210 sites defined as biotic properties were linked to environmental variables through stepwise multi-discriminant analysis (MDA), a multivariate statistical technique. There were 15 environmental variables used in the discriminant analysis, and the catchment area, stream width, altitude, water depth, velocity, BOD₅, SS, TN, TP, and conductivity were converted into \(\ln(x+1)\) to follow a normal distribution. Environmental variables (predictive variables) useful for distinguishing each TWINSPAN group and the discriminant functions to be applied to the model to obtain the probability of allocation for each TWINSPAN group at each site were determined based on Wilk’s lambda and significance \((p < 0.05)\) was derived from the discriminant analysis. The probability of allocation for each TWINSPAN group at each site, the probability of occurrence, and individual abundance/m² for each taxon were predicted according to the methods described in Moss et al. (1987) and Clarke et al. (2003), using predictive variables and discriminant functions derived from discriminant analysis. Because the main role of our biotic community predictive model is to suggest which organisms occur in a given environment, it was judged that the fit for the derived model would be more appropriate for similarity analysis than the method usually used in RIVPACS. Therefore, a similarity analysis was performed on the number of taxa (presence-absence data) and individual abundance/m² at each site using only the taxa predicted to match the observed taxa to confirm the probability that the predictive model will predict the observed taxa. For the number of taxa, Sørensen similarity (Sørensen 1948) and for individual abundance/m², Bray–Curtis dissimilarity (Bray and Curtis 1957) were applied, and the distribution of similarity and dissimilarity was derived from the 1,210 sites according to the stream types of Min et al. (2019). To confirm whether the additionally derived predictive model predicts the biotic community by reflecting the river continuum concept, the distribution of observed and predicted values by stream types was confirmed for the indicator of functional feeding groups based on shredders (SH), scrapers (SC), collector-filterers (CF), and collector-gatherers (CG), as shown by Equation (3) and used in Min et al. (2019).
Indicator value of functional feeding groups = \ln(CF + CG)/\ln(SH + SC)

= \log_{(SH+SC)}(CF + CG) \quad (3)

PCA and TWINSPLAN were analyzed using PC-ORD (MJM software, version 6.19, Gleneden Beach, OR, USA), and the Kruskal–Wallis test and stepwise multi-discriminant analysis were performed using PASW (for Windows 8, version 6.3 and SPSS Inc., Chicago, IL, USA).

3. Results

3.1. Overview of the environmental variables for the sampling sites grouped into watersheds

Before the analysis of the predictive model, the number of sites and environmental characteristics were summarized for each watershed (Table 2). In terms of catchment area, the sites at the Han and Nakdong Rivers, which had large watersheds, accounted for more than 50% of the total number of sites, whereas the smaller Seomjin and Yeongsan Rivers had a smaller number of sites. Based on the mean values for stream order, catchment area, and stream width, which represent stream size, the size of the sampled sites from

Table 2. Summary of the environmental variables assessed from 1,210 sampling sites by watershed.

| Environmental variables | Han River | Nakdong River | Geum River | Seomjin River | Yeongsan River |
|-------------------------|-----------|---------------|------------|---------------|---------------|
| Number of sites         | 367 (30.3)\(^a\) | 322 (26.6) | 254 (21.0) | 126 (10.4) | 141 (11.7) |
| Order                   | 3.7\(^b\)(1.5)\(^c\) | 4.7 (1.5) | 4.3 (1.3) | 3.8 (1.1) | 4.5 (1.4) |
| Range                   | 1–7       | 1–7          | 2–6        | 1–6          | 1–6          |
| Area (km²)              | 2076.5 (4541.5) | 5204.6 (7204.7) | 2427.1 (3549.8) | 544.1 (963.9) | 838.3 (1070.8) |
| Range                   | 2.9–23376.5 | 6.3–23433.1 | 3.9–9916.0 | 4.8–4914.3 | 4.4–3421.3 |
| Width (m)               | 184.0 (208.3) | 296.2 (276.8) | 371.0 (499.4) | 126.5 (82.3) | 328.0 (418.4) |
| Range                   | 8.0–1600.0 | 10.0–1550.0 | 19.8–3000.0 | 15.9–462.3 | 5.3–1900.0 |
| Altitude (m)            | 149.7 (177.4) | 79.3 (94.0) | 63.2 (63.5) | 132.5 (75.3) | 27.6 (26.7) |
| Range                   | 2.7–888.0 | 1.0–569.4 | 1.0–362.0 | 4.0–375.0 | 1.0–172.0 |
| Latitude                | 37.516 (0.362) | 35.889 (0.509) | 36.332 (0.327) | 35.188 (0.300) | 35.028 (0.310) |
| Range                   | 36.565–38.387 | 34.814–37.175 | 35.544–36.999 | 34.373–35.740 | 33.246–35.509 |
| Longitude               | 127.682 (0.604) | 128.569 (0.394) | 127.202 (0.332) | 127.250 (0.200) | 126.742 (0.166) |
| Range                   | 126.388–129.324 | 127.608–129.466 | 126.329–127.944 | 126.645–127.801 | 126.216–127.017 |
| Depth (cm)              | 37.4 (45.3) | 104.1 (219.3) | 48.6 (81.3) | 32.7 (14.0) | 71.3 (147.5) |
| Range                   | 12.3–461.9 | 14.7–1226.4 | 10.4–620.0 | 13.0–85.0 | 13.3–1138.3 |
| Velocity (cm/s)         | 39.9 (24.7) | 18.4 (17.9) | 33.3 (26.6) | 40.2 (42.2) | 29.5 (36.3) |
| Range                   | 0.0–123.8 | 0.0–107.7 | 0.0–123.6 | 0.0–216.7 | 0.1–211.7 |
| Diameter (qₙₚ)          | −3.6 (2.0) | −2.6 (2.2) | −2.0 (2.3) | −4.0 (1.8) | −2.0 (1.9) |
| Range                   | −8.2–1.5 | −7.1–1.5 | −6.8–1.5 | −6.9–0.2 | −7.0–1.5 |
| BOD₅ (mg/L)             | 1.9 (1.8) | 1.6 (0.7) | 2.5 (1.7) | 1.1 (0.4) | 3.1 (1.6) |
| Range                   | 0.2–13.5 | 0.4–5.1 | 0.5–12.6 | 0.4–3.9 | 0.4–7.5 |
| pH                      | 7.9 (0.4) | 7.9 (0.3) | 7.8 (0.3) | 7.6 (0.3) | 7.6 (0.3) |
| Range                   | 6.2–8.7 | 6.9–8.8 | 6.6–10.7 | 6.8–8.2 | 6.9–8.2 |
| SS (mg/L)               | 6.9 (6.6) | 8.3 (8.9) | 12.9 (9.5) | 6.3 (9.6) | 15.2 (10.5) |
| Range                   | 0.6–74.0 | 0.5–142.2 | 0.9–53.6 | 0.8–103.7 | 0.7–85.2 |
| TN (mg/L)               | 3.483 (2.043) | 2.677 (1.038) | 3.364 (1.592) | 1.723 (0.730) | 3.290 (2.107) |
| Range                   | 0.553–13.981 | 0.730–8.534 | 0.735–11.081 | 0.201–5.400 | 0.609–14.941 |
| TP (mg/L)               | 0.067 (0.088) | 0.046 (0.034) | 0.095 (0.096) | 0.036 (0.031) | 0.109 (0.062) |
| Range                   | 0.006–0.632 | 0.008–0.320 | 0.007–0.898 | 0.011–0.315 | 0.013–0.366 |
| Conductivity (μS/cm)    | 341.6 (839.2) | 595.2 (2749.2) | 335.5 (249.5) | 631.0 (3440.2) | 337.0 (266.7) |
| Range                   | 37.1–15699.7 | 43.3–29533.0 | 40.0–3137.7 | 39.2–30711.7 | 80.0–1627.1 |

\(^a\)Ratio of the number of sites to the total number of sites by stream types.
\(^b\)Mean.
\(^c\)Standard deviation.
largest to smallest were as follows: stream order; Nakdong, Yeongsan, Geum, Seomjin, and Han Rivers; catchment area: Nakdong, Geum, Han, Yeongsan, and Seomjin Rivers; and stream width: Geum, Yeongsan, Nakdong, Han, and Seomjin Rivers. This means that stream order, catchment area, and stream width can be used to represent stream size, but they showed slightly different results depending on their characteristics (Min et al. 2019). Based on the maximum values for the catchment area, rivers were ordered from the largest to the smallest as follows: Nakdong, Han, Geum, Seomjin, and Yeongsan Rivers. This means that there were many smaller sampling sites in the Han River compared to the Nakdong and Geum Rivers and that the Seomjin River also had many smaller sites than the Yeongsan River. The maximum of the stream order was 6th or 7th which was not largely different from the catchment area and stream width, which is believed to result from the large number of tributaries joining the mainstream of each watershed as there are many mountainous streams in the Republic of Korea.

For the altitudes corresponding to the geographical features, there were many high sites, as listed in decreasing order at the Han, Seomjin, Nakdong, Geum, and Yeongsan Rivers based on the mean value, and similar results were shown for their maximum values. Latitude and longitude depend on the location of the watersheds, and the Han and Geum Rivers had relatively high latitudes, whereas the Nakdong, Seomjin, and Yeongsan Rivers had low latitudes. Longitude was highest at the Nakdong River located on the right side of the Republic of Korea and lowest for the Yeongsan, Geum, and Seomjin Rivers located on the left, and intermediate for the Han River.

The water depth corresponding to the microhabitat showed a tendency to depend on the depth of the sites where the Ponar grab sampling was conducted, as the depth of the sites in order was the Nakdong, Yeongsan, Geum, Han, and Seomjin Rivers based on the mean and maximum values. Velocity was fastest in the Seomjin River and lowest in the Nakdong River, and the Han, Geum, and Yeongsan Rivers had different velocity rankings according to their mean and maximum values. The mean diameter of the rivers in order from highest to lowest was the Yeongsan, Geum, Nakdong, Han, and Seomjin Rivers (coarse-grained), based on their mean values, but the ranking according to the maximum and minimum values was different. This is because, unlike the water depth, velocity, and mean diameter showed similar trends to those of the catchment areas.

The BOD$_5$, SS, TN, and TP, which correspond to water quality, showed their highest overall tendencies at the Yeongsan River and their lowest at the Seomjin River, based on mean values. The SS and TP were the highest at the Yeongsan River but considering that BOD$_5$ and TN were the highest in the Hangang River. The overall water quality of the Yeongsan River was poor when compared to other watersheds. In addition, the overall water quality showed a good trend in the order of the Nakdong, Han, and Geum Rivers. The conductivity was highest in the Seomjin, Nakdong, Han, Yeongsan, and Geum Rivers sequentially, based on the mean values and considering that the overall changes in conductivity due to salinity were at the sites corresponding to the brackish regions and were highest in the Seomjin River and lowest in the Geumgang River.

### 3.2. Principle component analysis of the environmental variables

To confirm the relationship between each environmental variable, the cumulative variance up to Axis 3 for the PCA performed based on the adjusted environmental values was 70.0% (Axis 1, 40.2%; Axis 2, 20.4%; Axis 3, 9.4%), and the 15 variables were classified into three groups by Axes 1 and 2 (Figure 2; Table 3). Axis 1 first divided the 15 environmental variables into two groups (first group, variables for stream size, water depth, mean
diameter, and variables for water quality; second group, geographical feature, and velocity). Axis 2 then subdivided variables with negative correlation (first group, variables for stream size, water depth, and pH; second group, geographical feature, and velocity; third group, mean diameter, and variables for water quality except pH) with Axis 1. Axis 3 divided the water depth as a separate variable, but the variance was low at 9.4%.

For Pearson’s correlation analysis between each PCA axis and environmental variables, stream order ($r = -0.858$, $p < 0.01$), catchment area ($r = -0.843$, $p < 0.01$), stream width ($r = -0.873$, $p < 0.01$), water depth ($r = -0.624$, $p < 0.01$), mean diameter ($r = -0.676$, $p < 0.01$), BOD$_5$ ($r = -0.501$, $p < 0.01$), SS ($r = -0.628$, $p < 0.01$) showed a negative correlation with Axis 1, and altitude ($r = 0.648$, $p < 0.01$), and velocity ($r = 0.643$, $p < 0.01$) showed positive correlations. Furthermore, latitude ($r = 0.239$, $p < 0.01$), longitude ($r = 0.112$, $p < 0.01$), pH ($r = -0.126$, $p < 0.01$), TN ($r = -0.273$, $p < 0.01$), TP ($r = -0.225$, $p < 0.01$), and conductivity ($r = -0.421$, $p < 0.01$) showed negative correlations with Axis 1.

### Table 3: Results of the ordination by principal component analysis (PCA) of the 1,210 sampling sites (a) axis summary of PCA and (b) Pearson correlation coefficients ($r$) for the three axes with 15 environmental variables.

| Environmental variables | Axis 1 | Axis 2 | Axis 3 |
|-------------------------|--------|--------|--------|
| (a) Axis summary statistics for the PCA results’ | | | |
| Eigenvalue              | 0.198  | 0.100  | 0.046  |
| % variance explained    | 40.2   | 20.4   | 9.4    |
| Cumulative % variance explained | 40.2  | 60.6   | 70.0   |
| (b) Pearson correlation coefficient for 15 environmental variables | | | |
| Order                   | $-0.858^{**}$ | 0.400$^{**}$ | $-0.065^{*}$ |
| Area                    | $-0.843^{**}$ | 0.467$^{**}$ | $-0.134^{**}$ |
| Width                   | $-0.873^{**}$ | 0.240$^{**}$ | $-0.095^{**}$ |
| Altitude                | 0.648$^{**}$ | 0.414$^{**}$ | $-0.188^{**}$ |
| Latitude                | 0.239$^{**}$ | 0.156$^{**}$ | $-0.814^{**}$ |
| Longitude               | 0.112$^{**}$ | 0.694$^{**}$ | $-0.081^{**}$ |
| Depth                   | $-0.624^{**}$ | 0.200$^{**}$ | 0.085$^{**}$ |
| Velocity                | 0.643$^{**}$ | 0.011    | $-0.406^{**}$ |
| Diameter                | $-0.676^{**}$ | $-0.474^{**}$ | $-0.026^{**}$ |
| pH                      | $-0.126^{**}$ | 0.363$^{**}$ | $-0.187^{**}$ |
| BOD$_5$                 | $-0.501^{**}$ | $-0.668^{**}$ | $-0.275^{**}$ |
| SS                      | $-0.628^{**}$ | $-0.497^{**}$ | $-0.154^{**}$ |
| TN                      | $-0.273^{**}$ | $-0.511^{**}$ | $-0.557^{**}$ |
| TP                      | $-0.225^{**}$ | $-0.710^{**}$ | $-0.299^{**}$ |
| Conductivity            | $-0.421^{**}$ | $-0.396^{**}$ | $-0.310^{**}$ |

*p value < 0.05.  
**p value < 0.01.

Figure 2. Principal component analysis (PCA) for the 1,210 sites based on the adjusted values for the 15 environmental variables (a) Axis 1 versus Axis 2 and (b) Axis 2 versus Axis 3. White circles and red arrow lengths indicate the 1,210 sampling sites and the proportional relative importance of each environmental variable, respectively.
$p < 0.01$), TP ($r = -0.225, p < 0.01$), and conductivity ($r = -0.421, p < 0.01$) were statistically significant in relation to Axis 1, but the correlation was low. Axis 2 showed a negative correlation with longitude ($r = 0.694, p < 0.01$) and positive correlations with $\text{BOD}_5$ ($r = -0.668, p < 0.01$), TN ($r = -0.511, p < 0.01$), and TP ($r = -0.710, p < 0.01$), and overall, although statistically significant with other variables, the correlation was low. The PCA showed that overall, variables corresponding to stream size, geographical feature, and water quality all have the same characteristics, and water depth, velocity, and mean diameter changed in relation to stream size, geographical feature, and water quality, respectively.

### 3.3. Classification of the sites based on benthic macroinvertebrate communities

The benthic macroinvertebrates identified from the 1,210 sites included 317 taxa, and the TWINSPAN based on the adjusted abundance values for each taxon divided the 1,210 sites into two groups at division level 1 and a total of six groups from division levels 2 and 3 (Figure 3). The distribution of the sites for each group was confirmed in terms of the watersheds, and in general, Group 1 was found to be distributed in the Han, Geum, and Yeongsan Rivers; Group 2 in the Nakdong River; Group 3 in the Han, Nakdong, and Geum Rivers; and Groups 4–6 in the Han River. Although for Groups 4-6 the sites were mainly concentrated in the Han River, most of the sites for Group 6 were located at high latitudes in the east, whereas Groups 5 and 4, in that order, tended to spread to the lower latitudes and westward.

![Figure 3](image-url)  
**Figure 3.** Two-way indicator species analysis (TWINSPAN) of the 1,210 sampling sites based on the benthic macroinvertebrate communities using individual abundance/m$^2$. The level represents the division level with the TWINSPAN and the numbers in parentheses indicate the number of sites for each group.
When the distribution of the environmental variables for each group was confirmed, the stream size increased in the order of Groups 6, 5, 1, 4, 3, and 2 overall, based on the mean values (Figure 4). Altitude decreased in the order of Groups 6, 5, 4, 3, 2, and 1. The latitude decreased in the order of Groups 6, 5, 4, 1, 3, and 2, and the longitude in the order of Groups 6, 5, 4, 2, 3, and 1, similar to the altitude. The latitude and longitude decreased in the order of Groups 6 to 4, and the sites were generally distributed in the Han River. For latitude, Groups 1–3, whose sites were relatively evenly distributed at the Han, Nakdong, Geum, and Yeongsan Rivers compared to the other groups, showed a

Figure 4. Distribution patterns of each environmental variable for Two-Way Indicator Species Analysis (TWINSPAN) groups. (a) Stream order, (b) catchment area, (c) stream width, (d) altitude, (e) latitude, (f) longitude, (g) water depth, (h) velocity, (i) mean diameter of particle size in substratum, (j) pH, (k) 5-day biochemical oxygen demand, (l) suspended solid, (m) total nitrogen, (n) total phosphorus, and (o) conductivity. The boxes represent the 25th and 75th percentiles, and the whiskers indicate the 5th and 95th percentiles with standard deviations (error bar). The horizontal dotted lines and solid lines in each box represent the mean and median, respectively. All environmental variables show statistically significant differences \( p < 0.001 \) in each group according to Kruskal–Wallis test.
similar trend, whereas longitude decreased in the order of Groups 2, 3, and 1, which were identified at the eastern side of the Han River and on the Nakdong River, showing a relatively similar trend as for altitude. The water depth showed a distribution similar to that of the stream size and it was the deepest in Group 2, where there were many sites in the deep Nakdong River. The water depth of Group 2 was very deep at some Ponar grab sampling sites, so the mean was between the 75th and 95th percentiles of the box graphs. As the velocity decreased in the order of Groups 6, 5, 4, 3, 1, and 2, the distribution tendency was opposite to that of the water depth. The mean diameter decreased in the order of Groups 1 to 6 (increasing the proportion of coarse particles) and showed a distribution similar to that for the altitude. Although pH was the lowest in Group 1 and the largest in Group 2, there was little difference in the groups as the variables pH. BOD5, SS, TN, TP, and conductivity increased in the order of Groups 1 to 6, showing a tendency opposite to that for altitude. Conductivity had very high values at certain sites in each group due to the characteristics of the brackish water or high pollution; hence, the mean was between the 75th and 95th percentiles of the box graphs when compared to the other variables for water quality. All environmental variables showed statistically significant differences according to the Kruskal–Wallis test for each group.

3.4. Selection of predictive variables and functions for the predictive model

Each TWINSPAN group for the sites grouped by biotic characteristics was linked to environmental features through stepwise multi-discriminate analysis. Among the 15 candidate environmental variables used for the discriminant analysis, the discriminant variables selected according to the analysis were used as predictive variables and the derived discriminant functions were used as predictive equations. The discriminant analysis estimated that 14 of the 15 variables had a high discriminant power in distinguishing each

Table 4. Selection of 14 environmental variables that distinguish each TWINSPAN group of the 1,210 sampling sites, the relative importance of each variable, and the estimated five discriminant functions by stepwise multi-discriminant analysis.

| Step | Variables added at each step | Wilks’ lambda | F     | p-value |
|------|-------------------------------|---------------|-------|---------|
| 1    | Diameter                      | 0.480         | 260.8 | < 0.001 |
| 2    | Area                          | 0.331         | 177.4 | < 0.001 |
| 3    | Altitude                      | 0.278         | 130.4 | < 0.001 |
| 4    | Velocity                      | 0.245         | 105.4 | < 0.001 |
| 5    | TP                            | 0.214         | 91.6  | < 0.001 |
| 6    | Latitude                      | 0.198         | 79.8  | < 0.001 |
| 7    | pH                            | 0.180         | 72.5  | < 0.001 |
| 8    | Longitude                     | 0.167         | 66.2  | < 0.001 |
| 9    | Conductivity                  | 0.158         | 60.7  | < 0.001 |
| 10   | Depth                         | 0.150         | 56.3  | < 0.001 |
| 11   | SS                            | 0.145         | 52.0  | < 0.001 |
| 12   | BOD5                          | 0.141         | 48.4  | < 0.001 |
| 13   | Order                         | 0.139         | 45.0  | < 0.001 |
| 14   | TN                            | 0.137         | 42.0  | < 0.001 |

| Discriminant function | Wilks’ lambda | Chi-square | p-value |
|-----------------------|---------------|------------|---------|
| 1                     | 0.137         | 2385.3     | < 0.001 |
| 2                     | 0.430         | 1011.5     | < 0.001 |
| 3                     | 0.697         | 433.6      | < 0.001 |
| 4                     | 0.917         | 104.4      | < 0.001 |
| 5                     | 0.992         | 9.3        | 0.507   |

Wilks’ lambda = \frac{\text{Variance within groups}}{\text{Total variance} \times (\text{variance within groups} + \text{variance between groups})}
TWINSPAN group, and the stream width was not statistically significant and was excluded from the predictive variables (Table 4). The discriminant power to distinguish each TWINSPAN group among the selected environmental variables increased as Wilk’s lambda decreased as each variable was added in the order of mean diameter, catchment area, altitude, velocity, TP, latitude, pH, longitude, conductivity, water depth, SS, BOD$_5$, stream order, and TN. Based on these 14 environmental variables, a total of five discriminant functions were estimated to classify six TWINSPAN groups. Among them, discriminant functions 1 and 2 had a low Wilk’s lambda and were statistically significant, whereas Wilk’s lambda was high and not statistically significant in function 5. Discrimination functions 3 and 4 were statistically significant, but Wilk’s lambda was $\geq 0.6$. For this reason, only discriminant functions 1 and 2 were selected as predictive equations for the predictive model.

To confirm the environmental characteristics of the TWINSPAN groups according to discriminant analysis, the distribution of each group was confirmed using the centroid of the canonical discrimination functions 1 and 2, where Wilk’s lambda was relatively low (Figure 5). Each explanatory power for the canonical discrimination functions 1 and 2 accounted for 67.7% and 19.5% of the variance, respectively, showing the greatest amount of explanatory power in the analysis results. The relative importance of environmental variables was confirmed using a structure matrix, not a standardized canonical discriminant function coefficient, considering the multicollinearity between the variables. Although canonical discriminant function 1 also had the effect of conductivity, Groups were mainly classified according to the mean diameter, altitude, SS, BOD$_5$, and velocity. Consequently, Groups 3–6 with mean diameter were assembled according to high altitude and velocity, and low organic and inorganic pollution was located on the negative side of function 1, and Groups 1 and 2 showing opposite characteristics were located on the positive side. Although the canonical determination of function 2 also influenced longitude, the Groups were classified mainly according to the catchment area, TP, stream order, and water depth. Therefore, as the size was small, Groups 2, 4, 5, and 6 with shallow depths and low nutrient statistics were located on the positive side of function 2, and as the size was large, Groups 1 and 3, which had deep depths and high nutrient statuses, were located on the negative side.
3.5. Prediction of benthic macroinvertebrates based on 14 environmental variables

Based on 14 environmental predictive variables and two discriminant functions, a similarity analysis was performed for each site, and the distribution of the functional feeding groups was confirmed by stream type to further confirm that the predictive model reflected the river continuum concept of the rivers (Figure 6). To confirm the probability that the predictive model uses the observed taxa at each site, only the predicted taxa that matched the observed taxa were used in the similarity analysis. For the number of taxa, the similarity for each stream type in the Sørensen similarity for the observed and predicted taxa ranged from 0.60 to 0.72 based on the mean, and Type I, II, and IV with high altitude overall showed relatively higher similarity than III, V, and VI with low altitude.
The dissimilarity between the observed and predicted individual abundance/m² of the taxa with the Bray–Curtis dissimilarity was in the range of 0.46-0.56 based on the mean by stream type, and the dissimilarity was relatively low in Type II and IV regardless of stream size or altitude. Among the Type III, V, and VI (low land streams) sites, which had relatively low similarities when compared to the other stream types in the number of taxa, the sites that had particularly low similarity were generally due to Odonata, Coleoptera, Diptera, Trichoptera, and Malacostraca. In Type III, the sites of similarity were low, mainly due to Lestidae among the Odonata and Hydrophilidae among the Coleoptera. In Type V, the major taxa that lowered the similarity were Libellulidae and Gomphidae among the Odonata, Hydrophilidae, and Dytiscidae among the Coleoptera. Type VI showed low similarity due to the Libellulidae and Gomphidae among the Odonata, Hydrophilidae and Dytiscidae among the Coleoptera, and Unionidae among the Mollusca. Types I, III, V, and VI, which showed relatively high dissimilarity compared to the other stream types in terms of individual abundance/m², were derived from the large differences in individual abundance/m² of Baetidae, Heptaneniidae, Ephemerellidae, and Leptophlebiidae among the Ephemeroptera; Hydropsychidae among Trichoptera; Gammaridae and Atyidae among the Malacostraca; Chironomidae among Diptera; and Tubificidae among Annelida.

The distribution of the number of taxa and the individual abundance/m² of the functional feeding groups derived from the observed and predicted taxa increased in the order of Type I, II, IV, III, V, and VI, and altitude had a relatively larger effect on the distribution of the functional feeding groups when compared to size. Both the number of taxa and individual abundance/m² by stream types were in the range of the 25th and 75th percentiles of the box graphs for the functional feeding groups, and the predicted values were lower than the observed values. However, there were no significant differences in the distributions of functional feeding groups, except for Types V and VI in individual abundance/m² based on the mean. Therefore, the predictive model derived when the similarity analysis and the analysis of the distribution of functional feeding groups were synthesized reflected the changes in the benthic macroinvertebrates according to the river continuum concept, and at the same time, it is possible to determine the biotic community for the given environment appropriately.

4. Discussion

4.1. Appropriateness of used data (identification level, mesh size, and combination of sampling data)

The level of identification of the organism varies according to the purpose of the researches (Hawkins et al. 2000; Hargett et al. 2007). For example, in the United Kingdom (Clarke et al. 2011) and Australia (Smith et al. 1999), river monitoring was performed using a family level rather than a species level. Although the assessment was not as accurate as the species level, it could be said that the advantages of the family level (quick sample analysis and assessment, and the minimization of error due to misidentification) were used. The species level was also used in research to assess the effects of anthropogenic disturbances through a detailed survey of specific rivers and to clarify the differences (Hawkins et al. 2000). The purpose of the model is to identify organisms expected to inhabit the given environmental conditions, which may be more appropriate at the species level than at the genus level if focused more on aspects such as river restoration rather than monitoring and assessment. However, in the current Korean benthic...
macroinvertebrates, 1) some taxa have recently been classified (Jung et al. 2014; 2020), 2) the taxonomic list is continuously being revised (Jung et al. 2011; Park and Kong 2020), and 3) there were many organisms that have not yet been classified as species (mainly Diptera). In addition, although it was classified at the species level due to the limitations of morphological identification using a dissecting microscope, there were cases where the genus level is more suitable. This was mainly in the Mollusks (Park et al. 1997) and aquatic insects (Park and Kong 2020), where there were several imagoes but little morphological difference in the larvae. These taxa are currently identified at the genus level. For this reason, some taxa have not been accumulated sufficiently at the species level, it was judged that the current genus level would be appropriate.

In researches comparing the sampling methods or in the literature on protocols for sampling methods of benthic macroinvertebrates, it was generally recommended to use a mesh size of 0.5 mm or less (Carter and Resh 2001; Beatty et al. 2006; Couto et al. 2010; Cunha et al. 2019). This was because most of the loss of organisms in the sampling was large when a mesh size of 1.0 mm is used. As a smaller mesh size is used, the loss of organisms will be reduced and high-quality quantitative assessment will be possible (Beatty et al. 2006; Couto et al. 2010). However, using a small mesh size requires more time and effort for sample analysis than using a larger mesh size. Barba et al. (2010) suggested that the number of taxa, richness, and diversity, which are important metrics in our study, did not differ significantly in the sampling results between 0.5 mm and 1.0 mm mesh sizes. Beatty et al. (2006) emphasized the importance of mesh size suitable for the research purpose. That is, it will be useful to use a small mesh size for a detailed survey of a specific river, and to use a relatively large mesh size for nationwide river monitoring (Morin et al. 2004). Since the main purpose of the data used in our study is to monitor rivers throughout the Republic of Korea, it is judged that the guideline recommends the use of 1.0 mm. Medupin (2020) used a mesh size of 1.0 mm for benthic macroinvertebrate sampling as in our study. For the purpose of the predictive model, it may be more appropriate to use a mesh size as small as possible, but it was judged that this could supplement the use of 1.0 mm by using long-term cumulative data.

In the past, the United States assessed using different survey and sample analysis methods according to climate zones and specific states or regions, and these differences tended to depend on individual researchers’ tendencies (Carter and Resh 2001). After that, with the introduction of USEPA’s National Aquatic Resource Surveys, survey and sample analysis methods were standardized for each coast, lake, river, and wetland and this means that standardization is important for comparison and use of data. Correa-Araneda et al. (2021) compared the sampling methods of benthic macroinvertebrates using D-frame net (or hand net), litterbags, and core sampler, the number of taxa was higher in D-frame net and less in core sampler. If this is linked with our study, the D-frame net can be said to be similar to a Surber sampler that surveys a large area of the streambed surface, and the core sampler is similar to a Ponar grab that samples a small area to a specific depth of the streambed. This implies that the number of taxa collected when using the Surber sampler is larger than that of the Ponar grab. The difference in the number of taxa between the Surber sampler and the Ponar grab was considered to be due to 1) as the sampling area increases, the number of taxa increases (Kong and Kim 2015) and 2) the point sampled by Ponar grab was the center of non-wadable streams, which had a high velocity but deep water depth, and the habitat of various benthic macroinvertebrates was somewhat limited (Carter and Fend 2001). To use the same results for Surber sampler and Ponar grab as much as possible, individual abundance was converted to individual abundance/m², but this is limited to abundance, and
the difference in the number of taxa seems inevitable. In addition, the results of the Ponar grab sampled for a relatively small area show a greater increase in abundance compared to the Surber sampler in the process of conversion to individual abundance/m². Despite these differences, we used the data of the non-wadable stream to confirm the distribution of benthic macroinvertebrates in various environmental conditions and to reflect this in the model. Excluding the central data, the predictive ability of the model will be limited to less than 1 m in water depth.

4.2. Predictive variables for a predictive model of biotic community

A total of 15 environmental variables were considered for the predictive model, which were artificially classified into stream size, geographical feature, microhabitat, and water quality (Table 1), but PCA divided into three categories through Axis 1 and 2. The variables corresponding to the microhabitat belonged to the size, geographical feature, and water quality of the river, and water depth, velocity, and mean diameter had characteristics similar to the variables, respectively. In general, as the size of the river increases, so does the discharge and water depth and consequently, the water depth seems to have similar characteristics to the stream size. The main variable in determining the velocity is the slope and, because altitude has a similar characteristic to slope (Min et al. 2019), it seems that the velocity belongs to this geographical feature. The mean diameter was generally expected to belong to a geographical feature, as the decreases in altitude and velocity became finer, but it belonged to water quality. The composition of streambed substrates can affect biotic communities independently of water quality (Jun et al. 2016; Kong and Kim 2016), but they are related to each other according to the river continuum concept and show overlapping characteristics (Vannote et al. 1980). As an example, the Benthic Macroinvertebrates Index (BMI; Kong et al. 2018), a biotic assessment index developed based on BOD₅, and the Benthic Macroinvertebrates Streambed Index (BMSI; Kim and Kong 2019) to assess the state of streambed substrates based on data with minimal effects of water quality showed a high correlation in the Republic of Korea (Kim and Kong 2019).

Through discriminant analysis, 14 environmental variables to be used in the predictive model were selected, and the stream width was excluded. The relative importance of environmental variables classifying each TWINSPLAN group was the mean diameter, catchment area, altitude, velocity, TP, latitude, pH, longitude, conductivity, water depth, SS, BOD₅, stream order, and TN. Accordingly, benthic macroinvertebrates in the Republic of Korea changed by microhabitat, geographical feature, and stream size, implying that the physical variables took precedence over water quality. Stream size accompanies a change in the biotic community from upstream to downstream (Vannote et al. 1980). Stream order, catchment area, and stream width can all represent the stream size, but the characteristics of each variable were slightly different (Table 2). In the Republic of Korea, there is a large mountainous area, so there are many mountainous streams with low altitudes overall (Bae et al. 2003; Min et al., 2020). Even if the streams are the same size in terms of the catchment area, the stream order will be different depending on the number of inflow tributaries. Furthermore, in large-scale rivers, the stream order does not increase beyond a certain value even though the catchment area increases steadily. The stream width can represent the scale of the stream shape better than the stream order, but since it measured the fragmentary stream width of the sampling sites (e.g., length of a bridge within the site), not the width of a specific reach of the sites (NIER. 2019), it does not always increase, but may be constant or decrease as it moves downstream. This part may
be less consistent than stream order. The catchment area not only reflects the size and shape of the stream well but also increases as the size increases, and is less variable with disturbance or time than width or order (Min and Kong 2021). Therefore, it is believed that the catchment area has a large influence on the changes of biotic communities based first on the size variable followed by stream order a secondary variable.

The importance of geographical features such as altitude, latitude, and longitude in describing changes in benthic macroinvertebrate communities means that they mainly determine the distribution of biotic communities according to water temperature, although there may be complex effects along with several variables (Hawkins et al. 2000; Hargett et al. 2007). In particular, the importance of the water temperature was more pronounced as it had a great influence on explaining biotic community changes in the order of altitude, latitude, and longitude. Groups 4–6 had higher altitudes, latitude, and longitude (regional feature of East high and West low) overall than Groups 1–3, and in particular, altitude was one variable that showed consistent changes in Groups 1–6 compared to other variables. As such, it has been suggested in other previous research that geographical features in rivers affect changes in the benthic macroinvertebrate community (Sandin 2003; Min and Kong 2021).

Stream size and geographical feature can potentially affect the biotic community, whereas microhabitat can be said a direct factor in determining the distribution of biotic community. Therefore, the composition of the streambed, water depth, and velocity are variables that directly determine the benthic macroinvertebrate community. It was considered when identifying changes in the biotic community due to historic changes in the environmental variables (Sandin 2003; Li et al. 2012; Min and Kong 2021). The high proportion of fine particles in the streambed substrates means the streambed is mainly composed of sand particles, which provides a monotonous habitat or shelter for organisms, so even in a good water-quality environment, the habitat of organisms may be restricted (Xuehua et al. 2009; Jun et al. 2016). Conversely, coarse particles affect biodiversity and abundance as they provide a variety of microhabitats for benthic macroinvertebrates (Sandin 2003; Merz and Ochikubo Chan 2005). Accordingly, the mean diameter seems to have the largest influence on the change in the biotic community among the variables of the microhabitat. Velocity has been suggested in several investigations to be one of the main variables influencing the biotic community along with the streambed (Wright et al. 1984; Malmqvist and Maki 1994). The velocity affects the availability of food, and the amount of dissolved oxygen, when velocities are high, organisms are provided with more organic matter and dissolved oxygen in a shorter time (Oliveira and Nessimian 2010). As the water depth increases, velocities slow and the sedimentation can make the streambed finer, so water depth is considered to be a variable that determines the streambed along with the velocity, which was an important variable in predicting the biotic community (Hawkins et al. 2000; Clarke et al. 2011).

Among the water quality, TP can represent the trophic status of rivers, and an appropriate concentration can control the number of primary producers to generate a healthy and sustainable river process, but when TP is high and attached algae thrive in the streambed, the habitat of the benthic macroinvertebrates is limited (Dudley et al. 1986; Tonkin et al. 2014). In the absence of water pollution, pH can be changed by ecological processes (Hargreaves and Whitton 1976; Pinol and Avila 1992), but also by geology. pH can affect benthic macroinvertebrate communities depending on the amount of hydrogen ions independent of organic pollution or nutritional states (Kullberg 1992; Carrie et al. 2015; Sangpradub 2017; Tampo et al. 2021). Content of hydrogen in a water body may vary depending on chemical and biological variables (photosynthesis, respiration,
decomposition of organic matter) and characteristics of the parent material constituting the streambed’s geology (Berezina 2001). Researches on the relationship between pH and benthic macroinvertebrate mainly focus on 1) biotic communities in rivers acidified by anthropogenic disturbances such as mine wastewater (Gerhardt et al. 2004; Sangpradub 2017), 2) pH values under the influence of geology and distribution of mollusks (Evans and Ray 2010; Carrie et al. 2015), and 3) differences in biotic communities due to changes in pH based on sampling data of several or specific rivers (Kullberg 1992; Berezina 2001; Thomsen and Friberg 2002; Tampo et al. 2021). In this study, the pH of the entire sampling site was in the range of 6.2 to 10.7, so it is believed that there was no extreme effect of pH on the distribution of organisms such as the effect of mine wastewater (usually pH less than 5). The limestone zone has a higher concentration of magnesium and calcium ions than the granite zone, so the pH is also relatively higher, and consequently, the limestone zone is a more suitable habitat for mollusks than the granite zone (Evans and Ray 2010; Carrie et al. 2015). The geology of the Republic of Korea is mainly granite, with limestone occurring only partially in some areas (Yun and Moon 2009). However, Min et al. (2019) suggested that the differences in benthic macroinvertebrate communities between limestone and granite zones were due to differences in stream size or altitude rather than geological effects. Bae and Park (2020) confirmed that the important variables for the distribution of gastropods in the Republic of Korea were the composition of the substratum, organic pollution, and nutritional state, and the effect of pH was insignificant. Therefore, in our study, the main effect of pH on the distribution of benthic macroinvertebrate communities was believed to be the presence or absence of habitation or difference in individual abundance of specific organisms according to the range of pH for each stream (or site) that appeared according to various environmental characteristics. Conductivity can also be affected by geology such as pH (Min et al. 2019), but given that the sites used in our study were upstream of the rivers and the estuaries where seawater flows, it was considered that the changes in conductivity were due to salinity (Kong 2019). Accordingly, conductivity can be used more than other environmental variables to account for changes in certain taxa in specific environments, such as brackish water, as it reflects salinity. BOD₅ and SS are variables that can represent organic and inorganic pollution, respectively, and this can affect the habitat due to organism sensitivity (or tolerance) (Min and Kong 2020). Even when the physical variables suitable for habitation of specific benthic macroinvertebrates in the rivers are in place, if the organic pollution is high, the habitat of the highly sensitive taxa is restricted, and a community composed of highly tolerant taxa may prevail (Boon 1988). According to these characteristics, the tolerance value based on benthic macroinvertebrates due to water pollution from the past has been used as a tool to assess the biotic water quality of rivers (Resh and Jackson 1993; Karr 1999; Smith et al. 1999). TN is often used with TP as a variable for trophic status, but the background nitrogen concentration is high in the Republic of Korea due to the overall eutrophication of water bodies due to intensive land use and narrow land areas (Cho et al. 2012; Min and Kong 2020). For this reason, it was judged that TN had a lower influence on biotic communities than other water-quality variables. Considering that most of the growth limiting factors for algae in the Republic of Korea’s rivers and lakes are TP (Kong 2019), the effect of TN on the biotic community is expected to be greater for ammonia nitrogen, a toxic substance than organic nitrogen (Min et al., 2021; USEPA. 2013).

In this study, the main variables affecting the distribution of the biotic community may change gradually from upstream to downstream according to the river continuum concept but may change more greatly due to anthropogenic disturbances according to
land use. Land use affects the habitat of aquatic organisms due to bank erosion, sedimentation, inflow of nutrients and organic matter, or the inflow of toxic substances can have a direct and fatal effect on organisms (Jun et al. 2011; USEPA. 2013). In general, the effects of land use showed different biotic communities at similar stream sizes or altitudes to those in no impact cases (Kasangaki et al. 2008; Jun et al. 2011; Deborde et al. 2016). This is mainly related to the decrease of organisms with high sensitivity, but it has also been suggested that it is also related to the food sources available to organisms depending on the outcome of the impact (Guilpart et al. 2012). Some studies have shown that moderate disturbance limited the habitat of highly sensitive organisms, but also increased diversity by supplying nutrients and organic matter (Wagenhoff et al. 2011; Deborde et al. 2016). This was thought to be because aquatic organisms are generally very sensitive to disturbance or there are more taxa with moderate sensitivity than highly tolerant organisms (Dahal et al. 2007; Guilpart et al. 2012; Xu et al. 2014). However, an increase in diversity is not always positive because anthropogenic disturbances show differences from the biotic communities that appear in the natural streams. Land use around streams is an important factor in the distribution of biotic communities, but we did not have access to data on these, and it was judged that physical and chemical variables could comprehensively reflect the impact of land use.

4.3. Fitting the predictive model based on similarity

Statistical models derived based on the state of various streams make it possible to compare the taxa that are actually inhabiting, and the taxa expected to occur in each environment (Wright et al. 2000; Clarke et al. 2003). Statistical models such as RIVPACS describe the taxa expected to occur at sites based on environmental variables, as they are derived by standardizing empirical data, but to finally use the developed model, it is necessary to assess how accurately the model predicts the biotic community (Cox et al. 1997). As the main purpose of RIVPACS is to monitor and assess the ecological conditions of a wide range of rivers, the fit of the developed model has mostly used the standard deviation (accuracy) and coefficient of determination ($R^2$; precision) of $O/E$ calculated for reference sites (Hawkins et al. 2000; Clarke et al. 2003; Hargett et al. 2007). However, because our predictive model is not aiming to assess biotic aquatic ecosystems, but to suggest the biotic communities expected to inhabit a given environment, the fit of the derived model was confirmed using the similarity between the taxa that were observed and predicted based on environmental variables, not the RIVPACS model’s method. At this time, only the predicted taxa matching the organisms observed were used to check how high the probability was that the organisms observed were predicted. Regarding the probability of occurrence for the observed and predicted taxa by stream types, the Sørensen similarity regarding the number of taxa was $\geq 0.5$ based on mean. Whereas the Bray–Curtis dissimilarity for individual abundance/m$^2$ was $\leq 0.5$ in some types, and overall, these were low when compared to the similarity of the number of taxa. Some of the similarity regarding the number of taxa was low because the predictive model predicted some of the observed taxa with a low probability of occurrence. Considering that the number of species increases as the amount of sampling increases (Arrhenius 1921; Preston 1948; Kong and Kim 2015), it was judged that some taxa with lower similarity were sampled by chance during multiple sampling, so it is judged that the frequency of occurrence was low. A high degree of dissimilarity in individual abundance/m$^2$ meant that there was a large difference between the observed value and the predicted value for some taxa, and this was because unlike the number of taxa, the individual abundance/m$^2$
for either observed value or predicted value was very high. Mainly because the predicted individual abundance/m² was higher than that observed for some taxa, there were relatively more sites of high dissimilarity than those without, and most of the sites where the dissimilarity was very high showed that the observed individual abundance/m² of some taxa was very high when compared to the predicted individual abundance/m². Baetidae, Heptageniidae, Ephemerellidae, Leptophlebidae, and Hydropsychidae are the taxa found to have particularly high abundances when compared to other organisms in the microhabitats such as riffles where the velocity is fast and the streambed is coarse (Merz and Ochikubo Chan 2005; Horta et al. 2009; Gezie et al. 2020; Min and Kong 2020; Vázquez et al. 2020). The high dissimilarity was determined to be due to the main sampling points being riffle or run. In addition, there were cases where the dissimilarity was high due to the Chironomidae and Tubificidae. Although these taxa show a high abundance even in rivers with little disturbance, based on high tolerance, they tend to show a high abundance even in habitats with high pollution or fine substratum (Pinder and Brinkhurst 2000; Carew et al. 2007; Al-Shami et al. 2010). The fact that individual abundance/m² had a large deviation when compared to the number of taxa was also the same in the distribution of functional feeding groups by stream types. The individual abundance/m² of the functional feeding groups of Type V and VI showed a higher value and a larger deviation from the observed value than the predicted value compared to other river types, these were also caused by Baetidae, Hydropsychidae, Chironomidae, and Tubificidae, which are mainly collectors (filterers and gatherers). In terms of individual abundance/m², the large difference in similarity between the observed and predicted taxa and the small deviation in the distribution of the functional feeding groups from the predicted values converged the observed values of taxa to values corresponding to specific environmental variables. Similar to the central limit theorem, it was determined that this was due to the characteristics of the predictive model that suggest predictive values for taxa based on the converged numerical values. Considering these characteristics in RIVPACS, the individual abundance is replaced by the log value, and categories are classified according to range and use (Wright 1994; Paisley et al. 2014). In addition, the similarity of individual abundance/m² was low compared to the number of taxa whereas the large deviation in the distribution of functional feeding groups was related to the preference of taxa for the given environment compared to the number of taxa reflecting the taxa that occurred accidentally during sampling. Therefore, changes in individual abundance/m² also reflected the change in the biotic community that prefers the environment along with the frequency of appearance (Dufrene and Legendre 1997). For these characteristics, RIVPACS, which used only the presence or absence of occurrence when clustering reference sites in the past, uses the current individual abundance together (Wright et al. 2000; Clarke et al. 2003).

This study attempted to develop a predictive model that suggests the biotic community that is expected to inhabit the environment during the restoration of ecological rivers in the Republic of Korea using currently available data, but used taxa were genus level, and there were some cases where the similarity was low in the individual abundance/m². However, most of taxa had high similarity and the low similarity was only shown by some specific taxa, so it is likely to be used for river restoration in the future. The error that lowers the fit of the derived predictive model may occur because of 1) mismeasurement of environmental variables or misidentification of taxa (Clarke et al. 2003), 2) differences between sampled microhabitats such as riffle, run, and pool (Parsons and Norris 1996), and 3) high redundancy of used variables (Clarke et al. 2011) besides the characteristics of the used taxa (Allan and Castillo 2007). In general, as stream size increases, stream order, catchment area, and stream width increase, the altitude decreases slope...
becomes lower, and as water depth deeper, velocity is slower and streambed becomes finer at the sampling points. Therefore, in RIVPACS, some of the variables showing high redundancy with each other were excluded from among the various candidate for predictive environmental variables available (Wright et al. 2000). The United Kingdom and Australia, which use predictive models, are improving their models by finding environmental variables that have a higher correlation in predicting biotic community and reflecting the differences in microhabitats (Smith et al. 1999; Clarke et al. 2011). Therefore, our predictive model can be improved in the future by considering attempts in the United Kingdom and Australia based on species-level taxa. Furthermore, the development and application of predictive models for specific watersheds and rivers are also expected to serve as an opportunity to suggest more accurate biotic communities when restoring rivers.

Acknowledgments

The authors thank all the researchers who participated in the Nationwide Aquatic Ecological Monitoring Program.

Disclosure statement

The authors have no conflicts of interest to declare.

Funding

This work was supported by Kyonggi University’s Graduate Research Assistantship 2022

Notes on contributors

Jeong-ki Min received Ph.D. from Kyonggi University, Republic of Korea and is currently working for environmental impact assessment.

Hyoju Lee is currently a master’s course student at Kyonggi University, Republic of Korea. She is studying aquatic ecology and is particularly focused on benthic macroinvertebrates.

Dongsoo Kong was previously the chief of the Han River Environment Research Center at the National Institute of Environmental Research. He is currently a professor at Kyonggi University, Republic of Korea and studies water-quality management, freshwater ecology, and benthic macroinvertebrates.

ORCID

Jeong-Ki Min http://orcid.org/0000-0002-7674-4423
Hyoju Lee http://orcid.org/0000-0002-4093-3599
Dongsoo Kong http://orcid.org/0000-0001-8438-3229

References

Allan JD, Castillo MM. 2007. Stream ecology: structure and function of running waters. Dordrecht: Springer Science and Business Media.
Al-Shami SA, Rawi CSM, HassanAhmad A, Nor SAM. 2010. Distribution of Chironomidae (Insecta: Diptera) in polluted rivers of the Juru River Basin, Penang, Malaysia. J Environ Sci (China). 22(11): 1718–1727.
Arrhenius O. 1921. Species and area. J Ecol. 9(1):95–99.
Bae MJ, Park YS. 2020. Key determinants of freshwater gastropod diversity and distribution: the implications for conservation and management. Water. 12(7):1908.
Craig LS, Palmer MA, Richardson DC, Filoso S, Bernhardt ES, Bledsoe BP, Doyle MW, Groffman PM, Hassett BA, Kaushal SS, et al. 2008. Stream restoration strategies for reducing river nitrogen loads. Front Ecol Environ. 6(10):529–538.

Cramer ML. 2012. Stream habitat restoration guidelines. Olympia, Washington, USA: Co-published by the Washington Department of Fish and Wildlife, Natural Resources, Transportation and Ecology, Washington State Recreation, and Conservation Office Puget Sound Partnership, and the U.S. Fish and Wildlife Service.

Cuffney TF, Brightbill RA, May JT, Waite IR. 2010. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. Ecol Appl. 20(5):1384–1401.

Cummins KW. 1962. An evaluation of some techniques for the collection and analysis of benthic samples with special emphasis on lotic waters. Am Midl Nat. 67(2):477–504.

Cunha LLD, Alves MIB, Ono ER, Uieda VS. 2019. Methods of sampling benthic invertebrates in lotic habitats: a spatial and temporal evaluation. Acta Limnol Bras. 31:e4.

Dahal BM, Sitaula BK, Sharma S, Bajracharya RM. 2007. Effects of agricultural intensification on the quality of rivers in rural watersheds of Nepal. J Food Agric Environ. 5(1):341.

Deborde DDD, Hernandez MBM, Magbanua FS. 2016. Benthic macroinvertebrate community as an indicator of stream health: the effects of land use on stream benthic macroinvertebrates. Sci Diliman. 28(2):2–26.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard A-H, Soto D, Stiassny MLJ, et al. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev Camb Philos Soc. 81(2):163–182.

Dudgeon D. 2000. The ecology of tropical Asian rivers and streams in relation to biodiversity conservation. Annu Rev Ecol Syst. 31(1):239–263.

Dudley TL, Cooper SD, Hemphill N. 1986. Effects of macroalgae on a stream invertebrate community. J N Am Benthol Soc. 5(2):93–106.

Dufrène M, Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr. 67(3):345–366.

Evans RR, Ray SJ. 2010. Distribution and environmental influences on freshwater gastropods from lotic systems and springs in Pennsylvania, USA, with conservation recommendations. Am Malacol Bull. 28(1-2):135–150.

Gerhardt A, De Bisthoven LJ, Soares AMVM. 2004. Macroinvertebrate response to acid mine drainage: community metrics and on-line behavioural toxicity bioassay. Environ Pollut. 130(2):263–274.

Gezie A, Mulat WL, Anteneh W, Dejen E, Kloos H, Mereta ST. 2020. Habitat suitability modelling of benthic macroinvertebrate community in wetlands of Lake Tana Watershed, Northwest Ethiopia. Wetlands. 40(4):853–864.

Gørtz P. 1998. Effects of stream restoration on the macroinvertebrate community in the River Esrom, Denmark. Aquatic Conserv Mar Freshw Ecosyst. 8(1):115–130.

Guilpart A, Roussel JM, Aubin J, Caquet T, Marle M, Le Bris H. 2012. The use of benthic invertebrate community and water quality analyses to assess ecological consequences of fish farm effluents in rivers. Ecol Indic. 23:356–365.

Haapala A, Muotka T, Laasonen P. 2003. Distribution of benthic macroinvertebrates and leaf litter in relation to streamed retentivity: implications for headwater stream restoration. Boreal Environ Res. 8(1):19–30.

Hargrett EG, ZumBerge JR, Hawkins CP, Olson JR. 2007. Development of a RIVPACS-type predictive model for bioassessment of wadeable streams in Wyoming. Ecol Indic. 7(4):807–826.

Hargreaves JW, Whitten BA. 1976. Effect of pH on growth of acid stream algae. Br Phycol J. 11(3):215–223.

Hawkins CP, Norris RH, Hogue JN, Feminella JW. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol Appl. 10(5):1456–1477.2.0.CO;2]

Hill MO. 1979. TWINSPAN: a FORTRAN program of arranging multivariate data in an ordered two way table by classification of individual and attributes. Section Ecology System. Ithaca, New Your 14850, U.S.A: Cornell University.

Horta F, Santos H, Tavares L, Antunes M, Pinheiro P, Callisto M. 2009. Assessment of benthic macroinvertebrate habitat suitability in a tropical watershed. In Proceedings of the 7th International Symposium on Ecohydraulics. p. 170–179.

Jun YC, Kim NY, Kim SH, Park YS, Kong DS, Hwang SJ. 2016. Spatial distribution of benthic macroinvertebrate assemblages in relation to environmental variables in Korean Nationwide streams. Water. 8(1):27.

Jun YC, Kim NY, Kwon SJ, Han SC, Hwang IC, Park JH, Won DH, Byun MS, Kong HY, Lee JE, et al. 2011. Effects of land use on benthic macroinvertebrate communities: comparison of two mountain streams in Korea. Ann Limnol Int J Lim. 47(S1):S35–S49.
Jung SW, Jäch MA, Bae YJ. 2014. Review of the Korean Elmidae (Coleoptera: Dryopoidea) with descriptions of three new species. Aquat Insects. 36(2):93–124.

Jung SW, Jäch MA, Bae YJ. 2020. Review of the water penny beetles (Coleoptera: Psephenidae) of the Korean Peninsula based on morphology and mitochondrial cytochrome c oxidase subunit I gene sequences. J Asia Pa Biodivers. 13(1):13–23.

Jung SW, Lee DH, Ham SA, Huh JM, Hwang JM, Bae YJ. 2011. Revised checklist of the Korean aquatic insects. Bull Entomol Res. 27:37–51.

Karr JR. 1999. Defining and measuring river health. Freshw Biol. 41(2):221–234.

Kasangaki A, Chapman IJ, Balirwa J. 2008. Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. Freshwater Biol. 53(4):681–697.

Kim D, Kong D. 2019. Analysis of water quality factors for benthic macroinvertebrates streambed index (BMSI) improvement in Korea. J Korean Soc Water Environ. 35(6):539–549.

Kim JY, Kim PJ, Hwang SJ, Lee JK, Lee SW, Park CH, Moon JS, Kong D-S. 2017. Korean stream types based on benthic macroinvertebrate communities according to stream size and altitude. J Freshw Ecol. 32(1):741–759.

Kim M. 2007. Suggestions for ecological stream restoration. J Environ Impact Assess. 16(1):59–68.

Kim PJ, Kim JY, Son SH, Won DH, Kong D. 2016. Comparative analysis of benthic macroinvertebrate communities before and after the restoration project in the Kyoungan stream. J Korean Soc Water Environ. 32(1):15–22.

Koehn J, Brierley G, Cant B, Lucas A. 2001. River restoration framework. National Rivers Consortium. Land and Water Australia.

Kong D, Kim AR. 2015. Analysis of the relationship between number of species and survey area of benthic macroinvertebrates using weibull distribution function. J Korean Soc Water Environ. 31(2):142–150.

Kong D, Kim JY. 2016. Development of benthic macroinvertebrates streambeded index (BMSI) for bioassesssment of stream physical habitat. J Korean Soc Water Environ. 32(1):1–14.

Kong D. 2019. Statistical analysis on water quality characteristics of large lakes in Korea. J Korean Soc Water Environ. 35(2):165–180.

Kong DS, Son SH, Hwang SJ, Won DH, Kim MC, Park JH, Jeon TS, Lee JE, Kim JH, Kim JS, et al. 2018. Development of benthic macroinvertebrates index (BMI) for biological assessment on stream environment. J Korean Soc Water Environ. 34(2):183–201.

Kullberg A. 1992. Benthic macroinvertebrate community structure in 20 streams of varying pH and humic content. Environ Pollut. 78(1-3):103–106.

Kwon OK, Min DK, Lee JR, Lee JS, Je JG, Choe BL. 2001. Korean mollusks with color illustration. Seoul, Korea: Hanguel Publishing Company.

Lake PS, Bond N, Reich P. 2007. Linking ecological theory with stream restoration. Freshwater Biol. 52(4):597–615.

Lancaster J, Hildrew AG. 1993. Characterizing in-stream flow refugia. Can J Fish Aquat Sci. 50(8):1663–1675.

Lee SW, Hwang S-J, Lee J-K, Jung D-I, Park Y-J, Kim J-T. 2011. Overview and application of the National aquatic ecological monitoring program (NAEMP) in Korea. Ann Limnol Int J Limn. 47(S1):S3–S14.

Li F, Chung N, Bae MJ, Kwon YS, Park YS. 2012. Relationships between stream macroinvertebrates and environmental variables at multiple spatial scales. Freshw Biol. 57(10):2107–2124.

Louhi P, Mykrä H, Paavola R, Huusko A, Vehanen T, Mäki-Petäys A, Muotka T. 2011. Twenty years of stream restoration in Finland: little response by benthic macroinvertebrate communities. Ecol Appl. 21(6):1950–1961.

Malmqvist B, Maki M. 1994. Benthic macroinvertebrate assemblages in north Swedish streams: environmental relationships. Ecography. 17(1):9–16.

Malmqvist B, Rundle S. 2002. Threats to the running water ecosystems of the world. Envir Conserv. 29(2):134–153.

ME. 2014. Technical guidelines for ecological river restoration project. Indira Paryavaran Bhava: Ministry of Environment.

Medupin C. 2020. Spatial and temporal variation of benthic macroinvertebrate communities along an urban river in Greater Manchester, UK. Environ Monit Assess. 192(2):1–20.

Merz JE, Ochikubo Chan LK. 2005. Effects of gravel augmentation on macroinvertebrate assemblages in a regulated California river. River Res Appl. 21(1):61–74.

Miller SW, Budy P, Schmidt JC. 2010. Quantifying macroinvertebrate responses to instream habitat restoration: applications of meta-analyses to river restoration. Restor Ecol. 18(1):8–19.

Min J-K, Kim Y-J, Kong D-S. 2019. Spatial distribution patterns of benthic macroinvertebrate functional feeding groups by stream size and gradient in Republic of Korea. J Freshw Ecol. 34(1):715–738.
Min JK, Kong DS. 2020. Distribution patterns of benthic macroinvertebrate communities based on multispatial scale environmental variables in the river systems of Republic of Korea. J Freshw Ecol. 35(1):323–347.

Min JK, Kong DS. 2021. Development of a predictive model for benthic macroinvertebrates by using environmental variables for the biological assessment of Korean streams. J Freshw Ecol. 36(1):189–216.

Morin A, Stephenson J, Strike J, Solimini AG. 2004. Sieve retention probabilities of stream benthic invertebrates. J N Amer Benthol Soc. 23(2):383–391.

Moss D, Furse MT, Wright JF, Armitage PD. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biol. 17(1):41–52.

Nelson SM, Lieberman DM. 2002. The influence of flow and other environmental factors on benthic invertebrates in the Sacramento river, USA. Hydrobiologia. 489(1/3):117–129.

NIER. 2015. Korean Reach File (KRF) of water environment information system. National Institute of Environmental Research. [accessed 2021 June] http://211.114.21.35/GIS/?droneMode=Y&pMENU_NO=34.

NIER. 2017. Biomonitoring Survey and Assessment Manual. Announcement No. 2017-439. National Institute of Environmental Research.

NIER. 2019. Guidelines for stream/river ecosystem survey and health assessment methods - stream part -. Announcement No. 2019-52. National Institute of Environmental Research.

NIER. 2021. Partial revision of water pollution process test standards. National Institute of Environmental Research. No. 2021-93. [accessed 2022 January] https://www.law.go.kr/%ED%96%89%EC%A0%95%EA%B7%EC%9C%B9%EC%88%98%EC%A7%88%EC%98%A4%EC%97%BC%EA%B3%B5%EC%A0%95%EC%88%9C%ED%97%8A%EC%A8%80.

Oliveira ALHD, Nessimian JL. 2010. Spatial distribution and functional feeding groups of aquatic insect communities in Serra da Bocaina streams, Southeastern Brazil. Acta. 22(4):424–441.

Ormerod SJ. 2003. Restoration in applied ecology: editor’s introduction. J Appl Ecol. 40(1):44–50.

Paisley MF, Trigg DJ, Walley WJ. 2014. Revision of the biological monitoring working party (BMWP) score system: derivation of present-only and abundance-related scores from field data. River Res Appl. 30(7):887–904.

Palmer MA, Bernhardt ES, Allan JD, Lake PS, Alexander G, Brooks S, Carr J, Clayton S, Dahm CN, Follstad Shah J, et al. 2005. Standards for ecologically successful river restoration. J Appl Ecol. 42(2):208–217.

Park GM, Jeong KH, Jung Y. 1997. A comparative study of two species of Viviparidae (Mollusca: Prosobranchia): Cipangopaludina chinensis malleata and C. japonica in Korea. Korean J Malacol. 13(1):9–19.

Park SJ, Kong D. 2020. A checklist of Trichoptera (Insecta) of the Korean Peninsula. JSR. 9(3):288–323.

Parsons M, Norris R. 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. Freshw Biol. 36(2):419–434.

Pinder AM, Brinkhurst RO. 2000. A review of the Tubificidae (Annelida: Oligochaeta) from Australian inland waters. MMV. 58(1):39–75.

Piñol J, Avila A. 1992. Streamwater pH, alkalinity, pCO2 and discharge relationships in some forested Mediterranean catchments. J Hydrol. 131(1–4):205–225.

Poquet JM, Alba-Tercedor J, Punti T, del Mar Sánchez-Montoya M, Robles S, Álvarez M, Zamora-Muñoz C, Sáinz-Cantero CE, Vidal-Abarca MR, Suárez ML, et al. 2009. The MEDiterranean prediction and classification system (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. Hydrobiologia. 623(1):153–171.

Preston FW. 1948. The commonness and rarity of species. Ecol. 29(3):254–283.

Resh VH, Jackson JK. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York, USA. p. 195–223.

Rosenberg DM, Danks HV, Lehmkuhl DM. 1986. Importance of insects in environmental impact assessment. Environ Manage. 10(6):773–783.

Sandin L. 2003. Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. Ecography. 26(3):269–282.

Sangpradub N. 2017. Benthic macroinvertebrates in streams contaminated by acid mine drainage-A pilot study from Thailand. Philipp J Syst Biol. 11(1):71–77.

Schanze J, Olfert A, Tourbier JT, Gersdorff I, Schwager T. 2004. Urban river basin enhancement methods: existing urban river rehabilitation schemes. Dresden: Leibniz Institute of Ecological and Regional Development, Dresden (IOER).

Selego SM, Rose CL, Merovich GT, Welsh SA, Anderson JT. 2012. Community-level response of fishes and aquatic macroinvertebrates to stream restoration in a third-order tributary of the Potomac River, USA. Int J Ecol. 753634:9.
Selvakumar A, O’Connor TP, Struck SD. 2010. Role of stream restoration on improving benthic macroinvertebrates and In-stream water quality in an urban watershed: case study. J Environ Eng. 136(1):127–139.

Smith MJ, Kay WR, Edward DHD, Papas PJ, Richardson KJS, Simpson JC, Pinder AM, Cale DJ, Horwitz PHJ, Davis JA, et al. 1999. AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. Freshw Biol. 41(2):269–282.

Sorensen TA. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol Skar. 5:1–34.

Strahler AN. 1957. Quantitative analysis of watershed geomorphology. EOS. 38(6):913–920.

Strayer DL, Dudgeon D. 2010. Freshwater biodiversity conservation: recent progress and future challenges. J Am Benthol Soc. 29(1):344–358.

Strayer DL. 2006. Challenges for freshwater invertebrate conservation. J Am Benthol Soc. 25(2):271–287.2.0.CO;2

Tampo L, Kaboré I, Alhassan EH, Ouédra A, Bawa LM, Djaneye-Boundjou G. 2021. Benthic macroinvertebrates as ecological indicators: their sensitivity to the water quality and human disturbances in a tropical river. Front Water. 3:662–765.

Thomsen AG, Friberg N. 2002. Growth and emergence of the stonefly Leuctra nigra in coniferous forest streams with contrasting pH. Freshw Biol. 47(6):1159–1172.

Tikkkanen P, Laasonen P, Muotka T, Huhta A, Kuusela K. 1994. Short-term recovery of benthos following disturbance from stream habitat rehabilitation. Hydrobiologia. 273(2):121–130.

Tonkin JD, Death RG, Barquin J. 2014. Periphyton control on stream invertebrate diversity: Is periphyton architecture more important than biomass? Mar Freshwater Res. 65(9):818–829.

USEPA. 2013. Aquatic life ambient water quality criteria for ammonia—freshwater. Washington DC: United States Environmental Protection Agency.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. 1980. The river continuum concept. Can J Fish Aquat Sci. 37(1):130–137.

Vázquez RF, Vimos-Lojano D, Hampel H. 2020. Habitat suitability curves for freshwater macroinvertebrates of Tropical Andean Rivers. Water. 12(10):2703.

Wagenhoff A, Townsend CR, Phillips N, Matthaei CD. 2011. Subsidy-stress and multiple-stressor effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams and rivers. Freshw Biol. 56(9):1916–1936.

WFD. 2000. E.U. Directive 2000/60/EC of the European parliament and of the council establishing a framework for the community action in the field of water policy. Brussels: The European Parliament and the Council of the European Union.

Woo HS. 2009. River restoration models and review of river restoration guides. J Korean Soc Civ Eng. 57(12):19–28.

Wright JF, Moss D, Armitage PD, Furse MT. 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshwater Biol. 14(3):221–256.

Wright JF, Sutcliffe DW, Furse MT. 2000. Assessing the biological quality of fresh waters. RIVPACS and Other Techniques. Ambleside, Cumbria, UK: Freshwater Biological Association.

Wright JF. 1994. Development of RIVPACS in the UK and the value of the underlying data-base. Limnética. 10(1):15–31.

Xu M, Wang Z, Duan X, Pan B. 2014. Effects of pollution on macroinvertebrates and water quality bioassessment. Hydrobiologia. 729(1):247–259.

Xuelhua D, Zhaoyn W, Mengzhen X, Zhang K. 2009. Effect of streambed sediment on benthic ecology. Int J Sediment Res. 24(3):325–338.

Yochum SE. 2018. Guidance for stream restoration. US department of agriculture, forest service, national stream and aquatic ecology center. Technical Note TN-102.4. CO: Fort Collins.

Yoon IB. 1995. In an illustrative guidebook of aquatic insects in Korea. Jeonghaengsa Publishing Company: Seoul, Korea.

Yun CW, Moon HS. 2009. Classification of forest vegetation type and environmental properties in limestone area of. Korea J Agric Life Sci. 43(2):1–8.