Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assessing decoupled biological responses in Korean stream biomonitoring using a multitaxon Community Dynamics Index framework

Byeong-Hun Han
Hanyang University
In-Hwan Cho
Hanyang University
Ha-Kyung Kim
Hanyang University
Eun-A Hwang
Hanyang University
Su-Ok Hwang
Hanyang University
Baik-Ho Kim
tigerk@hanyang.ac.kr

Hanyang University

Research Article

Keywords: Hellinger distance, epilithic diatoms, benthic macroinvertebrates, fish assemblages, monsoon disturbance, ecological assessment

Posted Date: August 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10087614/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Assessing decoupled biological responses in Korean stream biomonitoring using a multi-taxon Community Dynamics Index framework

Byeong-Hun Han[1,2], In-Hwan Cho[1,3], Ha-Kyung Kim[1,4], Eun-A Hwang[1,5], Su-Ok Hwang[6], Baik-Ho Kim[1,6,7,*]

[1] Department of Environmental Science, Hanyang University, Seoul 04763, Republic of Korea
[2] Dongmoon ENT Co., Ltd., Guro-gu, Seoul 08377, Republic of Korea
[3] Migang Environment & Consulting, Anyang-si, Gyeonggi-do 14057, Republic of Korea
[4] National Institute of Environmental Research, Incheon 22689, Republic of Korea
[5] Han River Environment Research Center, Yangpyeong-gun, Gyeonggi-do 12585, Republic of Korea
[6] Institute for Natural Sciences, Hanyang University, Seoul 04763, Republic of Korea
[7] Department of Life Science, Hanyang University, Seoul 04763, Republic of Korea

Author email addresses:
Byeong-Hun Han: tttyy3@naver.com
In-Hwan Cho: crice5237@naver.com
Ha-Kyung Kim: hosang9022@naver.com
Eun-A Hwang: euna1602@naver.com
Su-Ok Hwang: spring3974@naver.com
Baik-Ho Kim: tigerk@hanyang.ac.kr

*Corresponding author: Baik-Ho Kim,

Department of Life Science, Hanyang University,

Seoul 04763, Republic of Korea.

Email: tigerk@hanyang.ac.kr

Abstract

Environmental monitoring programs increasingly require tools that compare biological change across multiple assemblages without treating all taxa as interchangeable. We applied a multi-taxon Community Dynamics Index (CDI) to national Korean stream biomonitoring records from 2011 to 2023. A strict tri-taxon core of 13,386 paired site-years was constructed in which epilithic diatoms, benthic macroinvertebrates, and fish were recorded at the same site, in the same year, and in both survey rounds. For each assemblage, Hellinger-CDI measured community displacement between survey rounds using Hellinger-transformed relative-abundance vectors. Mean CDI was highest for diatoms (0.988), intermediate for macroinvertebrates (0.903), and lowest for fish (0.636). Pairwise cross-assemblage synchrony was near zero, and split-half reliability correction and a structural random-pairing null model indicated that this decoupling was not explained by sampling noise, taxon richness, dominance, or the structural ceiling of Hellinger distance alone. Response classes showed that mixed/intermediate and single-assemblage pathways were more common than whole-ecosystem high-CDI reorganization. The framework identifies which biological compartment—biofilm surface, benthic habitat, or fish connectivity—expressed the strongest survey-round displacement. Multi-taxon CDI provides an assessment layer for stream biomonitoring, but should be interpreted as rainfall-linked community displacement rather than direct rainfall causality until paired environmental-driver matrices are integrated.

Keywords: Hellinger distance; epilithic diatoms; benthic macroinvertebrates; fish assemblages; monsoon disturbance; ecological assessment

Introduction

Environmental monitoring and assessment programs increasingly require tools that can distinguish biological condition, temporal change, response pathway, and uncertainty across multiple biological assemblages. In river networks, climate variability, hydrological pulses, land-use pressure, nutrient and sediment delivery, substrate stability, thermal exposure, and habitat connectivity jointly define the ecological space available to aquatic organisms (Resh et al., 1988; Poff et al., 1997; Bunn and Arthington, 2002; Allan, 2004; Dudgeon et al., 2006; Reid et al., 2019). Under such interacting pressures, no single biological group can be assumed to capture the full monitoring signal. Modern freshwater biomonitoring therefore increasingly relies on multiple assemblages and multiple lines of evidence to determine whether observed biological change reflects water-quality stress, benthic habitat alteration, flow disturbance, connectivity limitation, or broader ecosystem reorganization (Karr, 1981, 1999; Barbour et al., 1999; Hering et al., 2010; Birk et al., 2012; Feio et al., 2021; Simaika et al., 2024).

This temporal and pathway-oriented problem becomes especially important under hydrological variability. Flow regime and disturbance are central organizing forces in stream ecosystems, and climate change is altering water-cycle dynamics, flood and drought risks, and other water-related extremes with direct consequences for freshwater ecosystems and their services (Poff et al., 1997; Bunn and Arthington, 2002; Caretta et al., 2022; IPCC, 2022). In monsoon-influenced regions, much of the annual rainfall is concentrated into a short seasonal window, producing rapid changes in runoff, turbidity, conductivity, nutrient transport, fine-sediment movement, substrate exposure, hydraulic refugia, and habitat connectivity. Biological responses to this seasonal forcing are unlikely to be uniform across assemblages because organisms differ in life history, mobility, habitat coupling, generation time, and the temporal scale over which they integrate environmental exposure.

Korean streams provide a strong national-scale setting for evaluating this problem. Routine biomonitoring is conducted across major river-basin regions under a monsoon-influenced climatic and hydrological setting, allowing repeated biological observations to be interpreted within a seasonal disturbance context. Previous Korean studies showed that rainfall-linked ecological response is not controlled by rainfall amount alone, but also by antecedent water quality, land use, soil permeability, and pre-existing community composition. In the Hantangang River, monsoon rainfall altered water quality and epilithic diatom composition, and the largest Community Dynamics Index (CDI) response occurred in mid-polluted sites during the year with the highest rainfall (Cho et al., 2020). In four central-western streams, rainfall amount and frequency were linked to conductivity, nutrients, land use, soil characteristics, and CDI variation, with strong responses in the Sapgyocheon stream where urban land use and low-permeability soils were prominent (Hwang et al., 2023). These studies indicate that rainfall-linked community reorganization is a multi-pathway ecological process rather than a simple hydrological event.

Diatoms, benthic macroinvertebrates, and fish are expected to encode different components of this process. Epilithic diatoms are short-generation, substrate-associated biofilm organisms that respond rapidly to nutrient enrichment, organic pollution, conductivity, turbidity, light limitation, shear stress, and substrate-surface disturbance (Watanabe et al., 1986; Kelly and Whitton, 1995; Rimet and Bouchez, 2012). The Korean Multimetric Diatom Index advanced regional diatom assessment by integrating trophic, organic-pollution, motility, Achnanthes/Navicula-type, and Gomphonema-related information into a locally grounded biological integrity framework (Cho et al., 2019). Benthic macroinvertebrates integrate benthic habitat condition, organic matter, sediment, substrate stability, and flow refugia over broader ecological windows (Rosenberg and Resh, 1993; Bonada et al., 2006; Jun et al., 2012). Fish represent reach-scale habitat structure, mobility, connectivity, refuge use, and recolonization potential

(Karr, 1981; Jackson et al., 2001; Magoulick and Kobza, 2003). These differences make the three assemblages complementary rather than interchangeable monitoring endpoints.

CDI provides a temporal complement to conventional condition assessment. Whereas biological integrity indices describe ecological condition at a survey point, CDI quantifies the magnitude of compositional displacement between paired community states. Recent Korean work further showed that CDI can be integrated with nonlinear ordination, functional traits, and machine-learning diagnostics to interpret ecological instability in stream diatom assemblages (Lee et al., 2026). However, CDI has been applied mainly to diatoms, even though rainfall pulses and seasonal hydrological change affect entire stream ecosystems. A key unresolved question is therefore whether the temporal-displacement signal detected in epilithic biofilms is shared by benthic macroinvertebrates and fish under the same site-year contrast, or whether each assemblage records a distinct biological response pathway.

This study uses CDI as a shared temporal-displacement geometry to address an environmental monitoring and assessment question: do epilithic diatoms, benthic macroinvertebrates, and fish record the same survey-round community-dynamics signal, or do they encode decoupled biological responses under identical site-year contrasts? Using a strict tri-taxon matched national biomonitoring dataset from 2011 to 2023, we compared assemblage-specific Hellinger-CDI values for the three assemblages. Our objectives were to: calculate comparable Hellinger-CDI values while preserving the paired temporal-displacement logic of CDI; quantify assemblage-specific displacement magnitude and relative response strength; test whether diatom, macroinvertebrate, and fish CDI values are synchronized or decoupled under the same monitoring contrast; classify paired site-years into dataset-relative response pathways; and define the boundary between rainfall-linked survey-round displacement and direct causal attribution. We expected the three assemblages not to behave as interchangeable monitoring endpoints because they encode different biological pathways: rapid biofilm-scale response in diatoms, benthic-habitat and resource response in macroinvertebrates, and mobility- or connectivity-mediated response in fish. Accordingly, the study was designed as a matched multi-assemblage stream biomonitoring analysis, not as a site-specific causal rainfall-response model.

Materials and methods

Study area, monitoring frame, and methodological continuity

We used national Korean stream biomonitoring records collected from 2011 to 2023 across wadeable streams and river sites in the Han, Nakdong, Geum, and Yeongsan river-basin regions. These basins encompass forested headwaters, agricultural and urbanized mid-reaches, and downstream reaches affected by channel modification, hydrological regulation, habitat alteration, and pollution pressure. The national

monitoring design provides a large and internally consistent basis for evaluating assemblage-specific temporal displacement across multiple biological groups. However, the dataset is observational rather than experimental; therefore, all analyses were framed as monitoring and assessment analyses, not as manipulative tests of single-driver causality.

The analytical design directly builds on the Korean CDI research sequence developed in our previous studies. Earlier studies quantified rainfall-linked displacement of epilithic diatom communities and showed that rainfall response depended on rainfall amount, rainfall frequency, antecedent water quality, land use, soil permeability, and pre-existing community condition (Cho et al., 2020; Hwang et al., 2023). A subsequent UMAP-CDI study extended the framework by linking diatom instability to nonlinear ordination, functional traits, and machine-learning diagnostics (Lee et al., 2026). The present study retains the paired-displacement logic of those studies but expands the ecological scope from one biofilm assemblage to three biological compartments: epilithic diatoms, benthic macroinvertebrates, and fish. This design allows the analysis to ask whether the same survey-round contrast is recorded as a shared community-displacement signal or as assemblage-specific biological responses.

Paired sampling design and analytical scope

The analytical unit was a paired site-year, defined by a unique site code, one survey year, and two valid monitoring rounds. Within the national monitoring frame, Round 1 was treated as the spring pre-monsoon survey and Round 2 as the fall post-monsoon survey. This paired structure generally brackets the East Asian summer monsoon window and provides the temporal contrast used for CDI calculation. Each CDI value therefore represents the distance between two observed community states for the same assemblage, at the same site, and in the same year.

This paired design is ecologically meaningful but should not be confused with a manipulative rainfall experiment. Rainfall-linked interpretation is justified by the monsoon-season monitoring context and by previous Korean CDI studies that explicitly examined pre- and post-monsoon contrasts (Cho et al., 2020; Hwang et al., 2023). However, direct causal attribution requires site-specific linkage to precipitation windows, discharge or flow proxies, water chemistry, land use, soil permeability, riparian condition, channel form, habitat structure, and hydrological connectivity. The primary inference was therefore framed as rainfall-linked survey-round displacement, not as direct rainfall causality. This claim boundary was maintained throughout the analysis, interpretation, and response-class design.

Dataset matching, inclusion criteria, and reporting boundary

The raw monitoring database was retained in full. Unmatched records were not deleted from the source archive, and missing assemblage records were not replaced with zeros. Instead, a derived analysis database was constructed using a predefined matching hierarchy. For the primary tri-taxon core, a site-year was retained only when epilithic diatoms, benthic macroinvertebrates, and fish were all recorded at the same site, in the same year, and in both survey rounds. This strict rule yielded 13,386 paired site-years for the main analysis (Table 1; Fig. 1). Because all three assemblages were compared within identical spatial and temporal units, the design minimized the bias that would arise if assemblage-specific CDI values were estimated from different subsets of sites, years, or survey-round contrasts.

Records that did not satisfy the full tri-taxon rule were retained in the source database and used only for supplementary robustness logic where appropriate. Valid paired observations within one assemblage were assigned to taxon-specific full paired datasets, and valid paired observations for two assemblages at the same site-year were assigned to pairwise matched datasets. These broader datasets were not used for primary cross-assemblage inference. This separation is essential because CDI can be calculated only between two valid observed community states. A missing assemblage record represents an unavailable observation, not true biological zero abundance. Zero-imputation would artificially inflate apparent turnover and could create false cross-assemblage contrasts.

Table 1 provides the main-text dataset structure before and after strict tri-taxon matching. It is retained in the main text because the strict matching rule is the central evidence boundary of the study. The fuller distinction among the source archive, strict primary core, taxon-specific paired datasets, and pairwise matched datasets is documented in Supplementary Table S1; all main-text cross-assemblage claims remain restricted to the 13,386 paired site-years in the strict tri-taxon core

Table 1. Dataset structure before and after strict tri-taxon matching for the multi-taxon CDI framework.

Item	Diatom	Macro-invertebrate	Fish	Intersection/core	Main-text use
Raw taxon records	692,145	367,165	188,605	-	Original taxon-level source rows.
Unique site codes	3,038	3,037	3,038	3,036	Spatial basis for strict matching.
Site-year-round samples	27,677	27,642	27,443	27,192	Samples defined by year, round, and site code.
Paired site-years by assemblage	13,751	13,703	13,584	13,386	Strict tri-taxon core for primary inference.

Raw taxon-name richness before harmonization	788	748	326	-	Taxonomic harmonization before CDI calculation.

This table is intentionally data-forward. The fuller distinction among the source archive, strict primary core, taxon-specific paired datasets, and pairwise matched datasets is documented in Supplementary Table S1; all main-text cross-assemblage claims remain restricted to the 13,386 paired site-years in the strict tri-taxon core.

Biological assemblages and taxonomic harmonization

The input data for each assemblage consisted of long-format taxon records containing site code, survey year, monitoring round, taxon name, and abundance or standing-crop value. The term abundance/standing-crop is used because the raw database includes count- or abundance-type records rather than verified, comparable mass production across all assemblages. The analysis therefore focuses on relative community composition and temporal displacement, not absolute biomass or production. Epilithic diatoms were treated as biofilm-level primary-producer records, benthic macroinvertebrates as benthic habitat, sediment, substrate, and organic-matter pathway records, and fish as mobile consumer, refuge, and connectivity pathway records. This assemblage-level separation is consistent with the logic of multi-element biomonitoring, in which different biological groups represent different exposure windows, habitat compartments, and ecological processes.

Taxonomic harmonization was conducted independently within each assemblage before matrix construction. Leading and trailing spaces were removed, spelling variants were standardized where they clearly represented the same operational unit, uncertain identifications were retained as distinct operational categories where needed, and authorship strings attached to fish names were removed when they created duplicate records. Diatoms and fish were analyzed at the lowest reliable taxonomic level available in the monitoring database. Macroinvertebrates were analyzed using the operational taxonomic units provided in the national records because national monitoring datasets often contain a mixture of species-, genus-, and family-level identifications. No additional specimen-level re-identification was conducted for this analysis; the study instead emphasizes reproducible use of the harmonized monitoring database and makes the taxonomic resolution explicit.

Field sampling followed the standardized protocols of the Korean National Aquatic Ecological Monitoring Program (NAEMP), which has been applied consistently across the biological assemblages and monitoring rounds analyzed here (Lee et al., 2011). Epilithic diatoms were collected from submerged

hard substrates, primarily cobbles in riffle habitats, and preserved for taxonomic identification under light microscopy. Benthic macroinvertebrates were collected by standardized kick-net sampling across representative riffle, run, and pool microhabitats within each reach. Fish were surveyed by backpack electrofishing over standardized wadeable reaches. The database supports interpretation under a common national monitoring framework, but it does not allow every protocol version, field crew, or detection-effort detail to be reconstructed for every record. Therefore, among-assemblage CDI comparisons remain conditional on differences in sampling gear, detection window, organism mobility, habitat coupling, and spatial integration scale.

Community matrices and Hellinger-CDI calculation

For each assemblage, a sample-by-taxon matrix was constructed, with rows representing site-year-round samples and columns representing harmonized taxa. Within otherwise valid samples, taxa not recorded were coded as zero. If an entire assemblage was missing from a site-year-round, the sample was treated as missing rather than as a zero-valued community. Round-1 and round-2 matrices were aligned by site-year so that each CDI value represented displacement between two observed community states for the same assemblage, site, and year. Samples with zero total abundance were excluded because relative composition cannot be defined for an empty community vector.

Community matrices are sparse, high-dimensional, and strongly influenced by differences in total abundance among samples and assemblages. Therefore, the primary analysis used relative abundance followed by Hellinger transformation, a compositionally appropriate transformation for species-abundance matrices prior to Euclidean distance-based analysis (Legendre and Gallagher, 2001; Legendre and De Caceres, 2013). For assemblage g, site i, year y, survey round r, and taxon s, the original abundance or standing-crop value was denoted as $x_{i,y,r,s,g}$. Relative abundance was calculated as:

$$p_{i,y,r,s,g} = x_{i,y,r,s,g} / \Sigma_{s'} x_{i,y,r,s',g},$$

where $\Sigma_{s'} x_{i,y,r,s',g}$ is the total abundance or standing-crop value across all taxa s′ in the same assemblage g, site i, year y, and survey round r.

The Hellinger-transformed value was calculated as the square root of relative abundance:

$$h_{i,y,r,s,g} = \sqrt{p_{i,y,r,s,g}} = \sqrt{(x_{i,y,r,s,g} / \Sigma_{s'} x_{i,y,r,s',g})}.$$

For each assemblage g, site i, and year y, Hellinger-CDI was calculated as the Euclidean distance between the round-1 and round-2 Hellinger-transformed community vectors:

$$CDI_{i,y,g} = \sqrt{[\Sigma_s(h_{i,y,2,s,g} - h_{i,y,1,s,g})^2]}.$$

Here, $p_{i,y,r,s,g}$ represents the relative abundance of taxon s in a given site-year-round sample, $h_{i,y,r,s,g}$ represents its Hellinger-transformed value, and $CDI_{i,y,g}$ represents round-1 to round-2 compositional displacement. Algebraically, the Hellinger distance between two relative-abundance vectors p1 and p2 is equivalent to $\sqrt{(2 - 2 \cdot BC)}$, where $BC = \Sigma_s\sqrt{(p1_s \cdot p2_s)}$ is the Bhattacharyya coefficient (Bhattacharyya, 1943; Rao, 1995; Cha, 2007). Hellinger-CDI is bounded between 0, when two community vectors are identical, and $\sqrt{2} \approx 1.414$, when the two communities share no taxa. This common mathematical scale supports cross-assemblage comparison, but it does not make raw CDI magnitudes biologically identical across assemblages. Therefore, raw Hellinger-CDI was interpreted together with standardized CDI, reliability checks, structural-null comparisons, and explicit ecological interpretation.

Earlier CDI studies used PCA-based or other ordination-based distances, and the UMAP-CDI framework demonstrated the value of nonlinear representation for diagnosing instability within a single assemblage (Cho et al., 2020; Hwang et al., 2023; Lee et al., 2026). In contrast, direct comparison of UMAP distances across diatoms, macroinvertebrates, and fish would require separate assemblage-specific ordination spaces whose coordinate scales are not inherently equivalent. Hellinger-CDI was selected as the primary cross-assemblage geometry because it is deterministic, ordination-free, compositionally appropriate, and identically applicable to all three assemblages.

Cross-assemblage comparisons used both raw Hellinger-CDI and within-assemblage standardized CDI values. Raw CDI values describe the magnitude of round-to-round compositional displacement within each assemblage, whereas standardized CDI values support comparison of relative displacement strength among assemblages with different richness, sampling properties, and ecological response scales. This distinction preserves the paired-displacement logic of CDI while avoiding overinterpretation of absolute distance magnitudes as directly equivalent biological responses.

Cross-assemblage synchrony, decoupling, and response classes

Cross-assemblage synchrony was evaluated within the strict tri-taxon core by comparing site-year CDI values among epilithic diatoms, benthic macroinvertebrates, and fish. Pearson correlation quantified linear synchrony in displacement magnitude, whereas Spearman rank correlation tested whether site-years with high CDI in one assemblage also ranked high in another. Because the central ecological question was whether the three assemblages recorded a shared turnover gradient or decoupled biological responses, near-zero correlations were interpreted as evidence of weak synchrony rather than as failure of the CDI framework.

Because near-zero correlations could arise either from genuine ecological decoupling or from low reproducibility of assemblage-specific CDI, two split-half checks were applied to the tri-taxon core. The analysis used 50 random replications for each split-half diagnostic. A count-split design estimated the reliability ceiling associated with sampling noise by allocating observed counts to two pseudo-replicates and recalculating CDI for each half. Spearman correlations between parallel CDI series were averaged and corrected using the Spearman-Brown formula (Spearman, 1910; Brown, 1910). A species-split design then provided a more conservative internal-consistency diagnostic by partitioning taxa into disjoint subsets. Observed cross-assemblage Pearson correlations were disattenuated by the geometric mean of paired reliability values, following standard logic for correcting correlations affected by measurement error (Spearman, 1910). This split-sample logic is consistent with resampling approaches used to evaluate comparability and signal-to-noise structure in large-scale biomonitoring (Stoddard et al., 2008; Cao and Hawkins, 2011).

Cross-assemblage decoupling was quantified at the site-year level after within-assemblage standardization of CDI values. We used the range of standardized CDI among the three assemblages, dominant-response identity, and the cross-assemblage coefficient of variation to describe how strongly displacement was concentrated in one assemblage or distributed across multiple assemblages. These metrics follow the broader ecological logic that community turnover, multivariate dispersion, and beta-diversity structure can reveal ecological variability and instability (Anderson, 2001; Anderson et al., 2006; Baselga, 2010; Legendre and De Caceres, 2013).

Site-years were assigned to diagnostic response classes using empirical assemblage-specific thresholds. Within each assemblage, CDI values at or above the 67th percentile of the core distribution were classified as high, values at or below the 33rd percentile were classified as low, and the remaining values were classified as intermediate. Combinations of these states were translated into diagnostic response classes: diatom-sensitive early-warning response, macroinvertebrate-dominant habitat response, fish-dominant mobility/connectivity response, coupled biofilm-benthic response, coupled benthic-fish response, diatom-fish decoupled response, whole-ecosystem high-CDI reorganization, stable/low-CDI all-taxa response, and mixed/intermediate response.

These classes describe how temporal displacement was distributed among assemblages, not whether a site was ecologically good or bad. The empirical thresholds provide a reproducible, dataset-relative language for comparing biofilm-surface, benthic-habitat, fish-connectivity, coupled, and mixed response pathways under the declared Hellinger-CDI geometry. They should not be interpreted as biological-quality classes, impairment boundaries, restoration targets, or transferable regulatory cut-offs.

Sensitivity analyses, reproducibility, and causal-inference boundary

Sensitivity analyses tested whether the main conclusions depended on dataset filtering, transformation choice, distance metric, or ordination representation. Taxon-specific full paired datasets were compared with the strict tri-taxon core, and pairwise matched datasets were used for two-assemblage sensitivity checks. Primary Hellinger-CDI results were compared with relative Euclidean distance, log10(x + 1)-transformed abundance, Bray-Curtis dissimilarity, PCA/PCoA-based representations, and UMAP-based representations (Bray and Curtis, 1957; McInnes et al., 2018; Oksanen et al., 2022). Agreement was evaluated using Pearson correlation, Spearman rank correlation, response-class agreement, and stability of dominant-response identity.

A structural random-pairing null model tested whether the observed CDI magnitude ordering could be generated solely by differences in richness, dominance, sparsity, or the structural ceiling of Hellinger distance. The model follows null-model logic in community ecology, in which an observed pattern is compared against a distribution generated by deliberately breaking the structure of interest while preserving key marginal properties (Raup and Crick, 1979; Gotelli and Graves, 1996; Gotelli, 2000; Manly, 2007; Chase et al., 2011). Round-1 community vectors were kept in place, whereas Round-2 vectors were randomly reassigned within each assemblage. Hellinger-CDI was recalculated for 30 independent permutations, and the mean null CDI was compared with the observed mean CDI. The obs/null ratio, persistence index (1 − obs/null), and null-observed gap were used to summarize the degree of temporal identity retained relative to the disrupted-pairing expectation.

Reproducibility was organized around a derived paired site-year CDI table rather than redistribution of the restricted raw national monitoring database. The minimum reproducible table contains one row per valid site-year in the strict tri-taxon core, assemblage-specific Hellinger-CDI values, within-assemblage standardized CDI values, low/intermediate/high CDI states, response-class identity, year, basin region, and a filtering flag distinguishing primary, pairwise, and taxon-specific analyses. The Supplementary Information defines this derived-data schema and documents the split-half, structural-null, annual, basin-level, and distance-metric robustness outputs.

CDI measures temporal community displacement; it does not by itself indicate whether a change is beneficial, harmful, reversible, caused by rainfall, or driven by a specific environmental pathway. The study therefore separates measured survey-round displacement, rainfall-linked interpretation based on the monsoon-season monitoring context, and causal driver attribution. The third level requires direct linkage to rainfall windows, discharge or flow proxies, water quality, land use, soil permeability, habitat structure, riparian condition, and hydrological connectivity. This boundary prevents causal overstatement while

preserving the value of multi-taxon CDI as a reproducible diagnostic framework for comparing assemblage-specific biological responses in stream ecosystems.

Results

Dataset structure and tri-taxon matching

The national monitoring database provided a large multi-assemblage foundation for comparing temporal community displacement across Korean streams. After applying the strict tri-taxon matching rule, the final primary dataset contained 13,386 paired site-years (Table 1; Fig. 1), representing 96.58% of the union of site-years with paired observations in at least one assemblage. Taxon-specific full paired datasets and pairwise matched datasets were retained only for supplementary sensitivity analysis and were not used to infer primary cross-assemblage synchrony or decoupling. Thus, all main cross-assemblage results are based on the strict tri-taxon core in which diatoms, macroinvertebrates, and fish were observed under the same site-year and survey-round contrast.

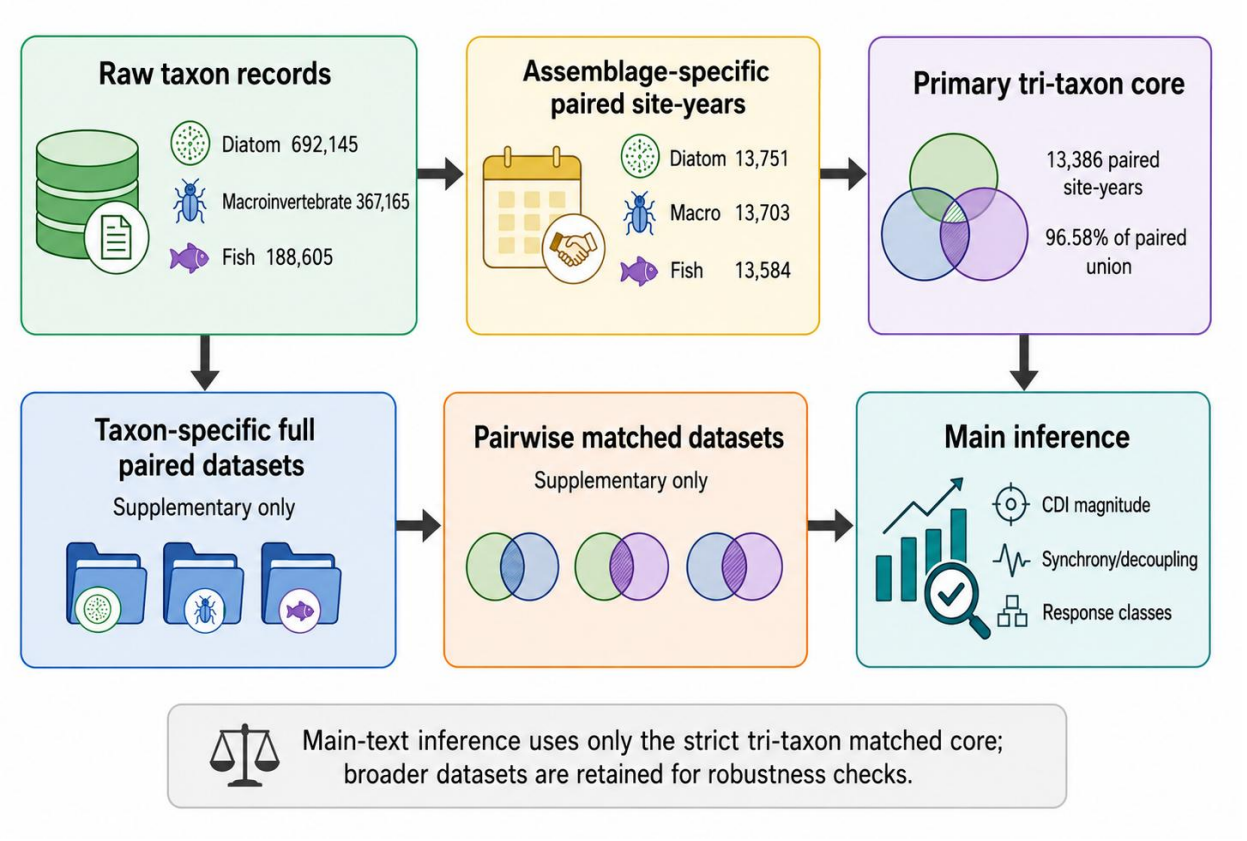


Fig. 1. Dataset construction and inference boundary for the multi-taxon CDI framework. Raw taxon records were organized into taxon-specific paired datasets, pairwise matched datasets, and the strict

primary tri-taxon core used for main-text inference. Broader matched datasets were retained for robustness checks only.

Assemblage-specific CDI magnitude

Hellinger-CDI values differed among the three assemblages (Fig. 2; Supplementary Table S2). Epilithic diatoms showed the highest mean CDI (0.988), followed by benthic macroinvertebrates (0.903), whereas fish showed substantially lower mean CDI (0.636). Expressed as a fraction of the theoretical maximum ($\sqrt{2} \approx 1.414$), these values correspond to 69.9%, 63.9%, and 45.0% of the maximum possible compositional displacement, respectively, indicating that all three assemblages routinely traversed a substantial fraction of compositional space between rounds but that the fraction was systematically smaller for fish. Median values followed the same ordering, with diatoms at 1.008, macroinvertebrates at 0.934, and fish at 0.612. These results indicate that survey-round community displacement was not equivalent among assemblages, even though all CDI values were calculated from the same paired site-years. This ordering supports the interpretation that the same monitoring contrast can be expressed with different displacement magnitudes across biofilm-surface, benthic-habitat, and fish-connectivity compartments. Detailed distributional statistics are provided in Supplementary Table S2.

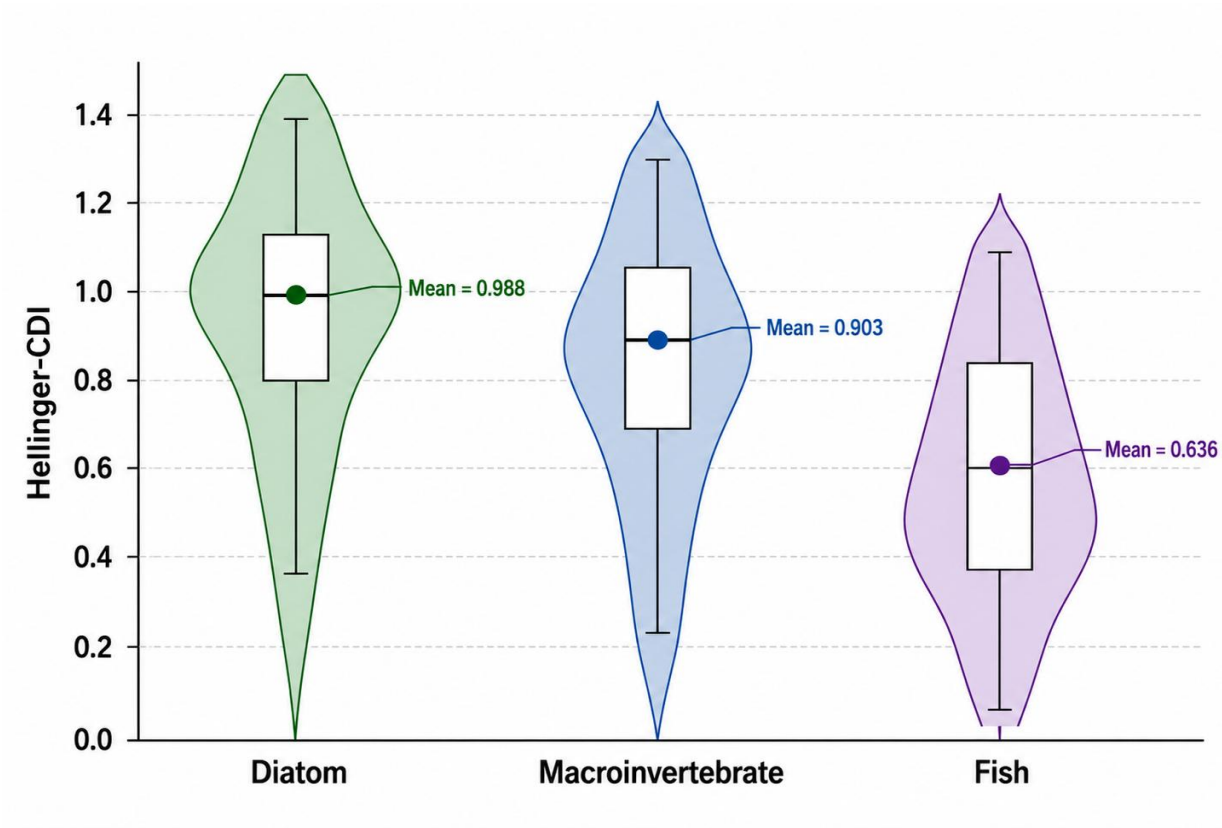
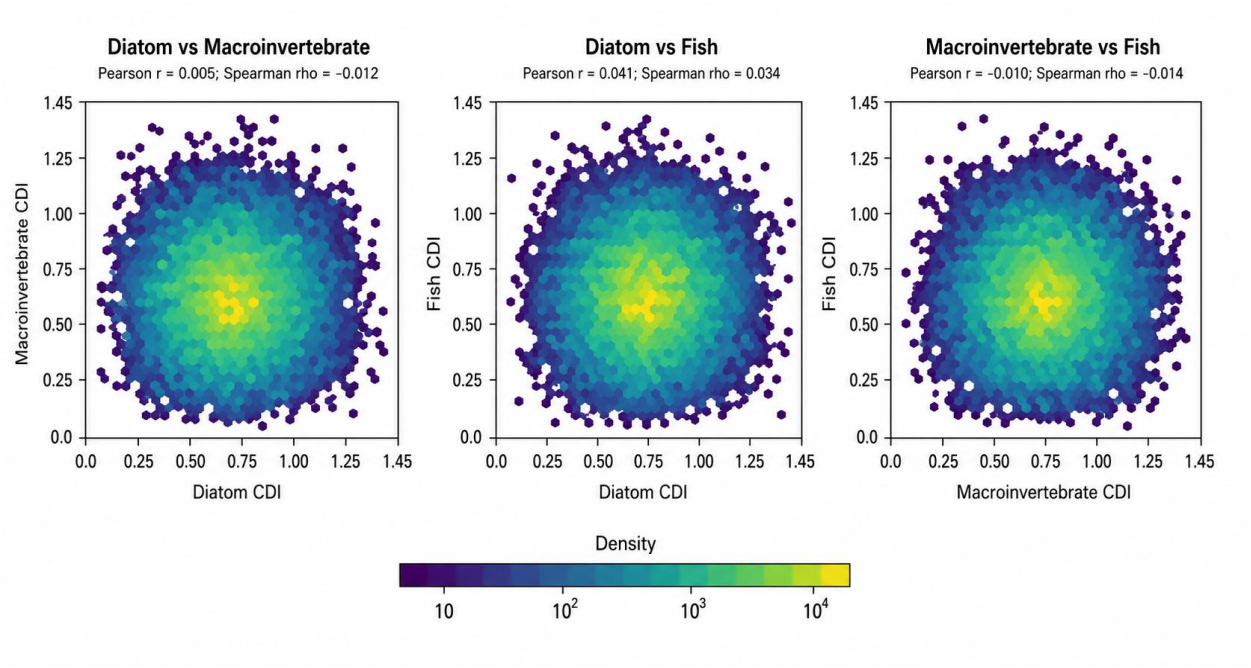


Fig. 2. Assemblage-specific Hellinger-CDI magnitude across the strict tri-taxon core. Violin and box plots show the distribution of paired survey-round community displacement for epilithic diatoms, benthic macroinvertebrates, and fish.

Cross-assemblage synchrony

Cross-assemblage synchrony was weak across all pairwise comparisons (Fig. 3; Supplementary Table S3). Diatom-CDI and macroinvertebrate-CDI were almost uncorrelated (Pearson $r = 0.005$; Spearman $\rho = -0.012$). Diatom-CDI and fish-CDI showed only a very weak positive association (Pearson $r = 0.041$; Spearman $\rho = 0.034$), and macroinvertebrate-CDI and fish-CDI were effectively uncorrelated (Pearson $r = -0.010$; Spearman $\rho = -0.014$). These near-zero relationships indicate that high displacement in one assemblage did not reliably predict high displacement in another assemblage. The primary result is therefore not a shared multi-taxon turnover gradient, but weak synchrony and decoupled assemblage-specific temporal displacement under the same matched monitoring contrast. Detailed synchrony values are presented in Supplementary Table S3.

The near-zero cross-assemblage correlations were not plausibly attributable to low count-split reliability of the assemblage-specific CDI series. Spearman-Brown corrected count-split reliability, which captures the reproducibility of Hellinger-CDI under binomial sampling-noise dilution, was 1.000 for diatoms, 0.986 for macroinvertebrates, and 0.939 for fish (Supplementary Table S3). The corresponding count-split reliability ceilings on the observable Pearson correlation between paired assemblage-specific CDI series were 0.993 (diatom ↔ macroinvertebrate), 0.969 (diatom ↔ fish), and 0.962 (macroinvertebrate ↔ fish). After Spearman-Brown disattenuation against these count-split ceilings, the cross-assemblage Pearson correlations remained essentially unchanged at 0.005, 0.042, and -0.011, respectively. A more conservative species-split internal-consistency check (Spearman-Brown corrected values of 0.559 for diatoms, 0.546 for macroinvertebrates, and 0.418 for fish) was also performed; even under this stricter diagnostic, the corrected cross-assemblage correlations remained small (≤ 0.085 in absolute value). Thus, even after generously correcting for count-split reliability, no pairwise correlation approached the magnitudes that would be expected of a shared multi-taxon turnover signal, supporting the interpretation that the observed decoupling reflects biological and scale-dependent structure rather than simple measurement noise.



391

392 **Fig. 3**. Cross-assemblage synchrony of Hellinger-CDI values. Density scatterplots compare paired site-
393 year CDI values among diatoms, macroinvertebrates, and fish. Pearson and Spearman correlations
394 indicate weak synchrony among assemblage-specific displacement signals.

395

396 **Response classes and pathways**

397 The multi-taxon CDI classification translated assemblage-specific displacement into diagnostic response
398 classes (Fig. 4; Supplementary Table S4). The largest class was the mixed or intermediate response
399 category (3,602 paired site-years; 26.9%), which captured cases without a clear high-CDI dominant
400 pathway. Among single-assemblage high-response classes, diatom-sensitive early-warning site-years
401 were most frequent (1,965; 14.7%), followed by macroinvertebrate-dominant habitat responses (1,947;
402 14.5%) and fish-dominant mobility/connectivity responses (1,926; 14.4%). Coupled two-assemblage
403 responses were less frequent, and whole-ecosystem high-CDI reorganization occurred in 545 paired site-
404 years (4.1%). Stable or low-CDI states across all three assemblages accounted for 443 paired site-years
405 (3.3%). The class distribution indicates that whole-ecosystem reorganization was not the dominant
406 pattern; instead, assemblage-specific or partially coupled displacement pathways were relatively common.
407 These classes should be interpreted as diagnostic summaries of displacement distribution, not as
408 ecological-quality grades or causal-driver categories.

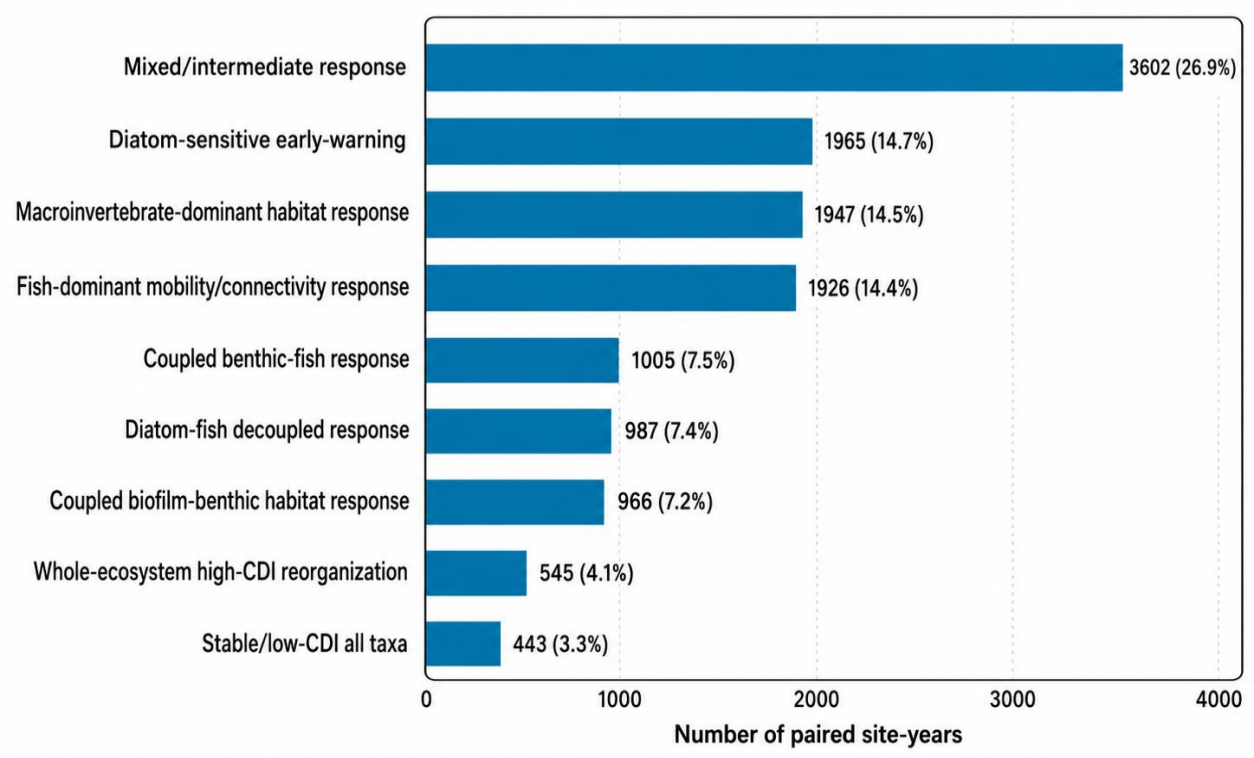


Fig. 4. Diagnostic response-class distribution in the multi-taxon CDI framework. Bars show the number and percentage of paired site-years assigned to mixed/intermediate, single-assemblage, coupled, whole-ecosystem high-CDI, and stable/low-CDI classes.

Temporal and basin-level patterns

Annual summaries showed that the relative ordering of assemblage-specific CDI was broadly maintained through time. However, individual years exhibited different magnitudes of displacement (Fig. 5; Supplementary Table S5). Diatom-CDI was highest in the early monitoring years and remained generally above fish CDI throughout the series. Macroinvertebrate-CDI showed a comparatively narrower range but increased in several mid-series years. Fish-CDI remained lower than the other two assemblages in all years, supporting the interpretation that fish assemblages recorded weaker round-to-round compositional displacement at the scale of this dataset. These annual patterns are descriptive and should not be interpreted as rainfall-driver effects until site-specific rainfall and environmental matrices are directly linked to the paired site-year CDI table.

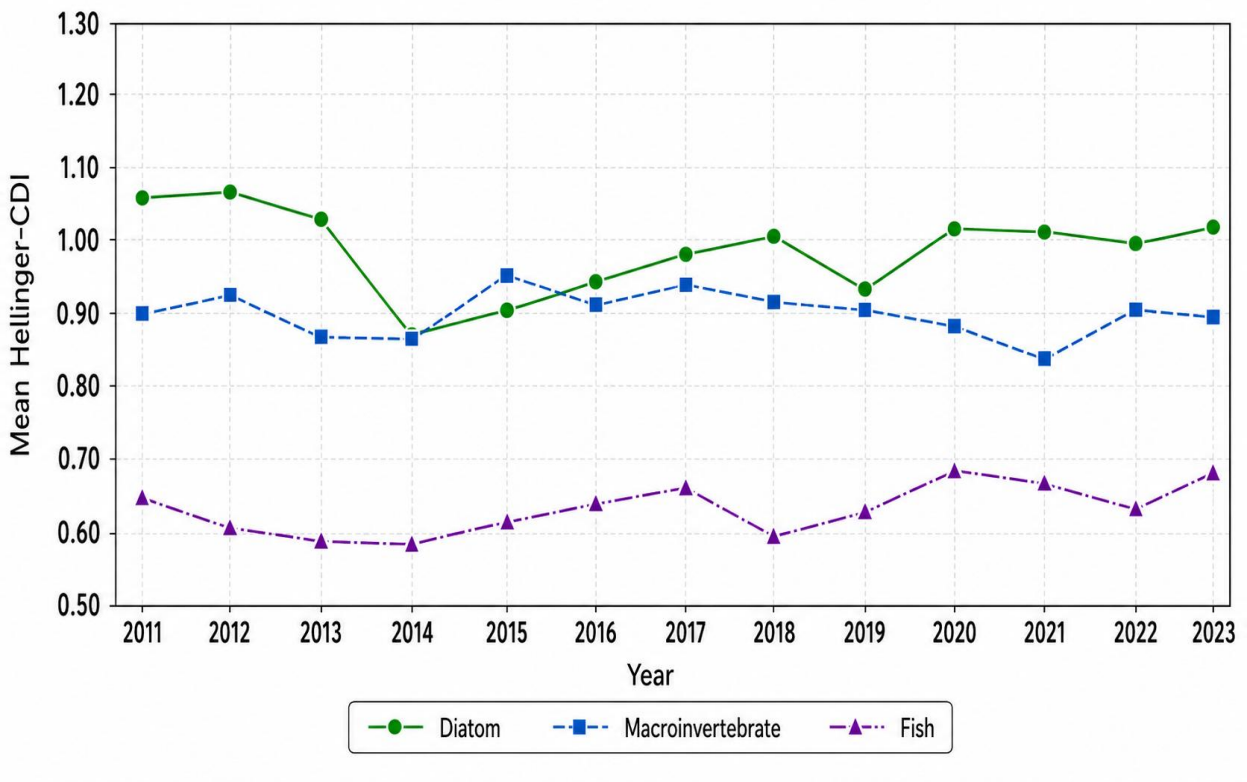


Fig. 5. Annual patterns of assemblage-specific Hellinger-CDI from 2011 to 2023. Mean CDI values are shown for diatoms, macroinvertebrates, and fish within the strict tri-taxon core.

Basin-level summaries revealed additional spatial structure in the multi-taxon CDI results (Fig. 6; Supplementary Table S6). The Geum River basin showed the highest mean diatom CDI (1.031) and the highest mean fish CDI (0.724), whereas the Nakdong River basin showed the highest mean macroinvertebrate CDI (0.971). Mean decoupling also varied among basins, with higher values in the Geum and Yeongsan basins than in the Han and Nakdong basins. These patterns indicate that assemblage-specific displacement and cross-assemblage decoupling were not spatially uniform. However, basin-level differences should be treated as descriptive until rainfall, water quality, land use, soil, habitat, and connectivity drivers are incorporated into the paired CDI table.

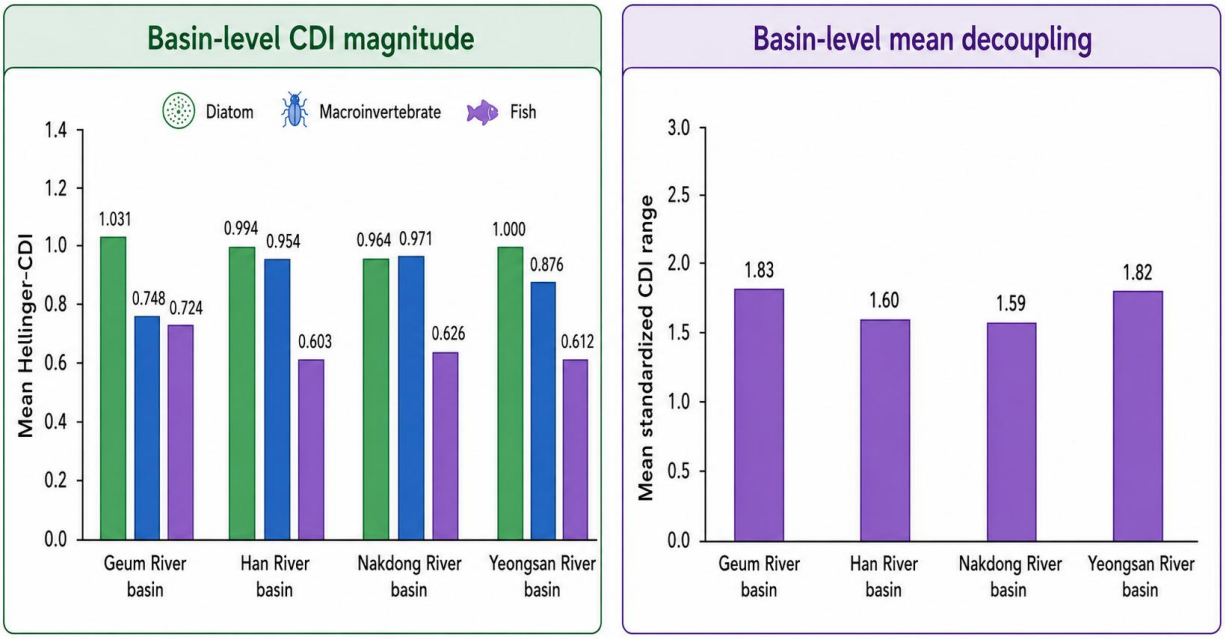


Fig. 6. Basin-level CDI magnitude and cross-assemblage decoupling. The left panel summarizes mean Hellinger-CDI by assemblage and basin, and the right panel summarizes basin-level mean standardized CDI range as a decoupling metric.

Robustness and results synthesis

The structural random-pairing null model further indicated that the among-assemblage ordering of observed Hellinger-CDI is unlikely to reflect a richness or dominance artifact alone. Despite large differences in mean richness per sample (25.1, 13.4, and 6.9 taxa for diatoms, macroinvertebrates, and fish, respectively) and within-sample dominance (mean maximum relative abundance of 0.355, 0.430, and 0.541), the null CDI values produced by random Round-2 reassignment were similar across the three assemblages (1.229 for diatoms, 1.238 for macroinvertebrates, and 1.262 for fish), and all three values approached the theoretical maximum of $\sqrt{2} \approx 1.414$. This pattern argues against the interpretation that the observed diatom > macroinvertebrate > fish ordering is simply a consequence of different richness or dominance structures. Expressed as obs/null ratios, observed displacement reached 80.4% of the null in diatoms, 72.9% in macroinvertebrates, and 50.4% in fish, and the corresponding persistence index (1 − obs/null) increased from 0.196 in diatoms through 0.271 in macroinvertebrates to 0.496 in fish. The absolute null-observed gap followed the same monotonic ordering (0.241, 0.335, and 0.626), so the rank order of temporal community identity retention was diatoms < macroinvertebrates < fish. This persistence

gradient is the inverse of the observed CDI magnitude ordering and is consistent with the expected ordering of generation time, mobility, and habitat coupling among the three assemblages.

Sensitivity analyses supported the main inference that assemblage-specific CDI responses were weakly synchronized across assemblages (Fig. 7; Supplementary Table S7). Hellinger-CDI showed broad rank agreement with relative Euclidean CDI and Bray-Curtis CDI, with mean Spearman correlations of 0.771 and 0.762, respectively, across assemblages. Strict core filtering did not alter mean CDI summaries relative to the full paired datasets, and the maximum absolute Pearson correlation among assemblage-specific Hellinger-CDI values remained at 0.041. Response-class assignments showed moderate sensitivity to the exact CDI metric; therefore, response classes should be interpreted as diagnostic categories under a declared metric geometry rather than fixed ecological states.

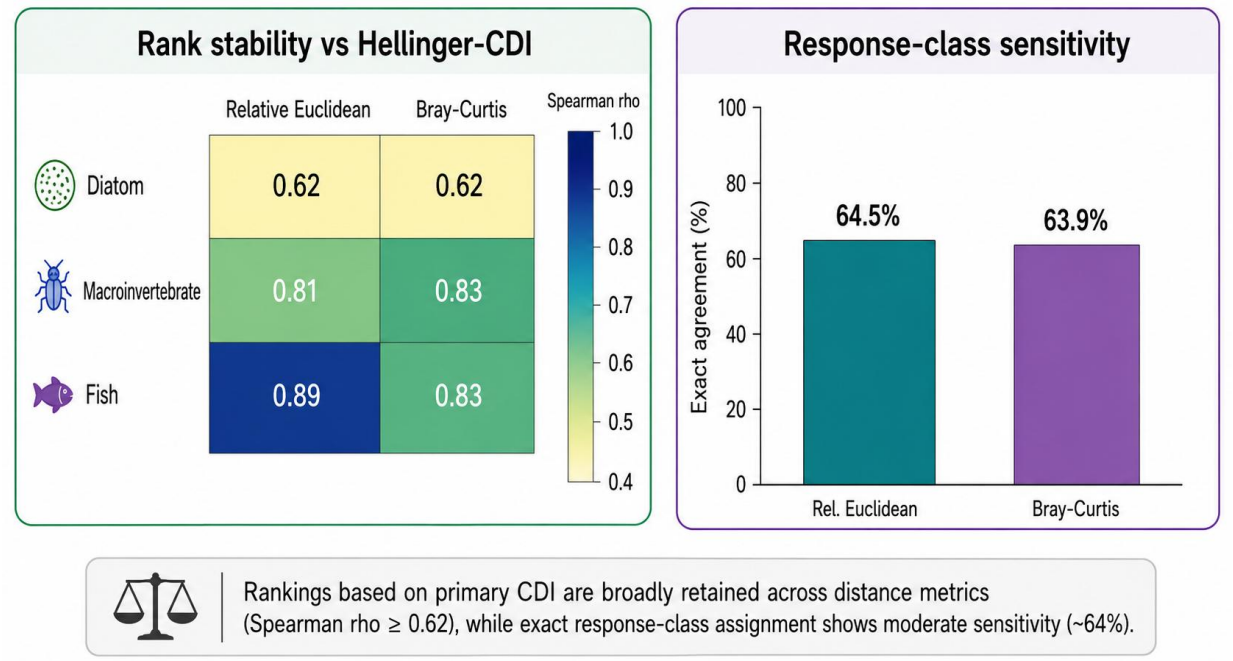


Fig. 7. Sensitivity of rank stability and response-class assignment to alternative distance metrics. Rank correlations compare relative Euclidean and Bray-Curtis results with primary Hellinger-CDI, while exact response-class agreement indicates moderate sensitivity of class membership.

Collectively, the Results section establishes four empirical findings. First, temporal community displacement differed among assemblages, with diatoms and macroinvertebrates showing larger average CDI than fish. Second, assemblage-specific CDI values were weakly synchronized even though all three biological groups were compared within the same strict site-year and survey-round structure, and this

weak synchrony was robust to Spearman-Brown disattenuation against high count-split reliability (≥ 0.939 for all assemblages), indicating that the near-zero pairwise correlations are not primarily a sampling-noise artifact. Third, a structural random-pairing null model produced similar null CDI values across the three assemblages (1.229–1.262) despite large differences in taxon richness and dominance, arguing against a richness- or dominance-driven artifact and revealing a persistence gradient (1 − obs/null = 0.196, 0.271, and 0.496 for diatoms, macroinvertebrates, and fish, respectively) that is consistent with increasing generation time, mobility, and habitat coupling. Fourth, the response-class framework provides a reproducible way of translating multi-taxon CDI combinations into diagnostic ecological pathways while preserving the important boundary that CDI measures displacement magnitude, not direction, ecological quality, or causal mechanism.

Discussion

Multi-taxon CDI as a temporal assessment layer

This study extends the Community Dynamics Index (CDI) from an assemblage-specific measure of temporal community displacement into a multi-taxon assessment layer for stream biomonitoring. The central design strength is the strict tri-taxon matched core of 13,386 paired site-years, in which epilithic diatoms, benthic macroinvertebrates, and fish were compared under the same site-year and survey-round contrasts. This matching rule is important because apparent differences among assemblages can otherwise reflect mismatched sites, years, sampling windows, or data completeness rather than true ecological divergence. By restricting the main inference to matched observations, the study treats multi-assemblage comparison as a controlled monitoring question rather than as a broad compilation of independent biological datasets.

The framework complements conventional biological assessment rather than replacing it. Biological integrity and multimetric indices remain essential because they summarize ecological condition, tolerance structure, and management status at a given survey point (Karr, 1981, 1999; Barbour et al., 1999; Stoddard et al., 2006; Hering et al., 2010; Birk et al., 2012; Feio et al., 2021). In Korea, the Korean Multimetric Diatom Index advanced stream assessment by integrating trophic, organic-pollution, motility, Achnanthes/Navicula-type, and Gomphonema-related information into a regionally grounded diatom index (Cho et al., 2019). However, condition indices do not directly ask whether communities move between repeated surveys, whether different assemblages move together, or whether temporal displacement is concentrated in one biological compartment. CDI addresses this temporal question. Static indices describe ecological condition; Hellinger-CDI describes survey-round displacement.

This study also extends our previous CDI work. Earlier rainfall-diatom studies used CDI to quantify pre-post monsoon displacement of epilithic diatom communities and showed that rainfall-linked response depended on rainfall amount and frequency, antecedent water quality, land use, soil permeability, and pre-existing community composition (Cho et al., 2020; Hwang et al., 2023). A later UMAP-CDI framework combined nonlinear ordination, functional traits, and machine-learning diagnostics to interpret ecological instability in Korean stream diatoms (Lee et al., 2026). The present study moves beyond the diatom-only question. It asks whether the temporal-displacement signal detected in biofilms is shared by benthic macroinvertebrates and fish under identical site-year contrasts. This shift from single-assemblage instability to multi-assemblage synchrony and decoupling is the main conceptual contribution of the manuscript.

Decoupled responses are ecological information, not analytical noise

The strongest empirical result is weak synchrony among assemblage-specific CDI values. Diatoms and macroinvertebrates showed larger average Hellinger-CDI values than fish, yet pairwise correlations among the three CDI series were close to zero. Thus, a site-year with high diatom displacement was not reliably a site-year with high macroinvertebrate or fish displacement. This finding should not be interpreted as failure of multi-taxon biomonitoring. It is the ecological signal that multi-assemblage monitoring is designed to detect. If all assemblages responded identically, one biological group could serve as a universal proxy. The near-zero synchrony observed here indicates that epilithic diatoms, benthic macroinvertebrates, and fish are not interchangeable endpoints of a single shared turnover gradient.

The decoupling is biologically plausible because the three assemblages integrate environmental change through different habitats, time scales, and mechanisms. Epilithic diatoms are attached biofilm organisms with short generation times and direct exposure to the substrate-water interface. They can respond rapidly to nutrients, conductivity, turbidity, light availability, shear stress, fine-sediment deposition, and substrate-surface disturbance (Kelly and Whitton, 1995; Rimet and Bouchez, 2012; Cho et al., 2019, 2020; Hwang et al., 2023). Benthic macroinvertebrates integrate benthic habitat, substrate stability, organic matter, sediment deposition, and hydraulic refugia over broader ecological windows (Rosenberg and Resh, 1993; Bonada et al., 2006; Jun et al., 2012). Fish respond at still broader reach scales through mobility, refuge use, recolonization, connectivity, and life-history filtering (Karr, 1981; Jackson et al., 2001; Magoulick and Kobza, 2003). The same monsoon-season contrast can therefore appear as strong biofilm reorganization, benthic-habitat displacement, or fish-connectivity response depending on which ecological pathway is most strongly expressed.

The reliability and structural-null analyses support this interpretation. Count-split reliability was high, and disattenuation against reliability ceilings did not convert the near-zero correlations into meaningful synchrony. The structural random-pairing null model also produced similar null CDI values across assemblages despite large differences in richness and dominance. These checks argue against the interpretation that the observed diatom > macroinvertebrate > fish CDI ordering is merely a sampling-noise effect or a richness/dominance artifact. Instead, the null-observed gap suggests a persistence gradient: diatoms retained the least temporal identity relative to the random-pairing ceiling, macroinvertebrates were intermediate, and fish retained the most. This ordering is consistent with expected differences in generation time, mobility, habitat coupling, and detection scale.

Response classes as diagnostic pathways

The response-class framework translates multi-taxon CDI combinations into diagnostic assessment pathways. This is useful because monitoring interpretation often requires knowing where to look next. A diatom-sensitive response directs attention to biofilm-surface processes, short-term water-quality pulses, turbidity, light limitation, nutrient delivery, and substrate-surface disturbance. A macroinvertebrate-dominant response directs attention to benthic habitat, fine sediment, organic matter, substrate stability, and hydraulic refugia. A fish-dominant response directs attention to connectivity, barriers, refuge availability, floodplain access, reach-scale habitat structure, and mobile-community redistribution. Coupled classes suggest propagation across compartments, whereas whole-ecosystem high-CDI cases identify site-years in which displacement was expressed simultaneously across biofilm, benthos, and fish.

The empirical distribution of response classes reinforces the decoupling result. Mixed/intermediate and single-assemblage high-response pathways were more frequent than whole-ecosystem high-CDI reorganization. This means that synchronized ecosystem-wide displacement was relatively rare, whereas compartment-specific or partially coupled displacement was common. For monitoring programs, this prevents two common errors: a strong diatom response should not automatically be interpreted as whole-ecosystem change, and a weak fish response should not be interpreted as absence of ecological reorganization. Different assemblages detect different pressure pathways and temporal windows, which is why modern freshwater assessment increasingly relies on multiple biological quality elements rather than a single biological endpoint (Karr, 1981; Barbour et al., 1999; Hering et al., 2010; Birk et al., 2012; Feio et al., 2021; Simaika et al., 2024).

These response classes are not ecological-quality grades, impairment thresholds, or restoration targets. They are empirical diagnostic categories generated under a declared Hellinger-CDI geometry and assemblage-specific 33rd/67th percentile thresholds. Their value is to organize evidence and indicate

which ecological compartment should be examined next. In this sense, the response-class framework functions as a decision-support layer, not a regulatory classification system.

Monitoring value under monsoon-influenced hydrology

The framework is especially relevant to Korean and other monsoon-influenced stream networks. Monsoon rainfall can dilute pollutants, mobilize nutrients, resuspend fine sediment, scour or bury biofilms, restructure benthic habitats, modify hydraulic refugia, and alter connectivity. Previous Korean rainfall-diatom studies showed that ecological response to rainfall is not controlled by rainfall amount alone, but depends on antecedent water quality, land use, soil permeability, and initial community composition (Cho et al., 2020; Hwang et al., 2023). The present study extends this logic to a national multi-assemblage setting. Rather than asking whether rainfall caused a single uniform response, it identifies which biological compartment expressed survey-round displacement most strongly.

This perspective is consistent with broader river-ecology theory. Flow regime, disturbance, sediment movement, habitat structure, and connectivity are central organizing forces in stream ecosystems (Resh et al., 1988; Poff et al., 1997; Bunn and Arthington, 2002; Allan, 2004). Freshwater biodiversity is increasingly affected by interacting pressures, including hydrological alteration, land-use change, pollution, habitat fragmentation, invasive species, and climate variability (Dudgeon et al., 2006; Reid et al., 2019; Tickner et al., 2020). Large-scale evidence also suggests that freshwater biodiversity recovery can stagnate even after earlier improvement, emphasizing the need for monitoring systems that distinguish water-quality recovery, habitat limitation, connectivity constraints, and biological response failure (Haase et al., 2023). Multi-taxon CDI contributes to this need by making temporal displacement and compartment-specific response pathways visible.

Interpretation boundaries and future integration

Several boundaries should remain explicit. CDI is a magnitude metric. It does not determine whether community change represents degradation, recovery, seasonal turnover, recolonization, homogenization, or stochastic replacement. Directional ecological interpretation requires taxonomic composition, trait structure, tolerance values, richness, dominance, biological integrity scores, water chemistry, habitat variables, and hydrological data. Raw CDI magnitudes are mathematically comparable under Hellinger geometry, but they are not biologically identical across assemblages because diatom scraping, macroinvertebrate kick-netting, and fish electrofishing differ in sampling gear, spatial grain, detectability, organism mobility, and effective temporal integration. Strict matching, Hellinger transformation, within-assemblage standardization, reliability correction, and null models reduce these issues, but they do not remove all biological and protocol-scale differences.

The structural random-pairing null model is informative but bounded. It shows that the observed CDI ordering is unlikely to be explained only by within-sample richness and dominance, but it does not preserve basin-specific spatial autocorrelation or protocol-specific detection windows. A natural next test would use within-basin or stratified null models by basin, stream order, or habitat type. Similarly, response-class thresholds are dataset-relative and should not be exported as universal thresholds without recalibration.

The next stage is driver-linked and trait-mediated CDI modeling. Paired site-years should be connected to antecedent rainfall windows, discharge or flow proxies, turbidity, nutrients, conductivity, organic pollution, land use, soil permeability, riparian condition, channel form, substrate composition, habitat structure, barriers, and hydrological connectivity. Trait information should also be integrated: diatom growth form and motility, macroinvertebrate habit, feeding group and voltinism, and fish mobility, trophic guild and reproductive strategy could convert CDI from a taxonomic displacement metric into a more mechanistic ecosystem-response framework. Operationally, CDI should be used beside conventional biological integrity scores and environmental measurements. The most defensible next stage is integration, not replacement: CDI should help monitoring programs distinguish ecological condition, temporal change, biological pathway, and causal uncertainty under increasingly variable monsoon climates.

Conclusions

This study tested a multi-taxon CDI framework for Korean stream biomonitoring using a strict tri-taxon matched core dataset comprising 13,386 paired site-years. By applying a common Hellinger-based displacement geometry to epilithic diatoms, benthic macroinvertebrates, and fish, the study shows that temporal community displacement can be compared across assemblages when sampling units are carefully matched and cross-assemblage interpretation remains biologically cautious.

The main empirical result is decoupling. Diatom, macroinvertebrate, and fish CDI values were weakly synchronized, indicating that the three assemblages did not record a single shared turnover gradient. Instead, they expressed assemblage-specific biological responses to the same monitoring contrast: rapid biofilm-scale response, benthic-habitat and resource response, and mobility- or connectivity-mediated response. This decoupling is unlikely to be a simple measurement artifact. Spearman-Brown disattenuation against count-split reliability values of 0.939-1.000 left the pairwise correlations near zero, and a structural random-pairing null model produced similar null CDI values across assemblages (1.229-1.262), arguing against an explanation based solely on taxon richness or dominance. The null-observed

gap instead defined a persistence gradient that increased from diatoms to macroinvertebrates to fish, consistent with increasing generation time, mobility, habitat coupling, and detection scale.

The response-class framework converts assemblage-specific CDI combinations into diagnostic assessment pathways, including diatom-sensitive, macroinvertebrate-dominant, fish-dominant, coupled, mixed, stable, and whole-ecosystem high-CDI responses. These classes are not ecological-quality grades, impairment thresholds, or causal-driver categories. Rather, they provide a practical structure for identifying which ecological compartment and driver pathway should be examined next.

Overall, multi-taxon CDI complements conventional biological integrity assessment by adding a temporal, pathway-oriented layer to freshwater community analysis. Static indices describe ecological condition, whereas multi-taxon CDI describes temporal displacement and whether that displacement is biofilm-specific, benthic-habitat-mediated, fish-mediated, coupled, mixed, or ecosystem-wide. The present results should therefore be interpreted as rainfall-linked survey-round displacement rather than direct rainfall causality. The next step is to link paired site-years to rainfall windows, discharge or flow proxies, water quality, land use, soil permeability, channel structure, riparian condition, habitat connectivity, and assemblage-specific traits. Such integration would allow multi-taxon CDI to support more mechanistic ecological interpretation and management prioritization under increasingly variable monsoon climates.

Data availability

The derived summary outputs supporting the main figures, supplementary tables, and robustness analyses are provided in the Supplementary Information and the accompanying derived-output workbooks. These materials include the reviewable source values needed to verify the reported Hellinger-CDI distributions, cross-assemblage synchrony statistics, response-class frequencies, annual and basin-level summaries, and metric-sensitivity results. The raw national biomonitoring records are subject to program data-use restrictions and are not redistributed in this submission. Access to the original monitoring database should be requested from the responsible Korean monitoring authorities.

References

Allan, J.D., 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. 35, 257-284. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122.

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32-46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

662 Anderson, M.J., Ellingsen, K.E., McArdle, B.H., 2006. Multivariate dispersion as a measure of beta
663 diversity. Ecol. Lett. 9, 683-693. https://doi.org/10.1111/j.1461-0248.2006.00926.x.

664 Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid Bioassessment Protocols for Use
665 in Streams and Wadeable Rivers, 2nd ed. U.S. Environmental Protection Agency, Washington, DC.

666 Baselga, A., 2010. Partitioning the turnover and nestedness components of beta diversity. Glob. Ecol.
667 Biogeogr. 19, 134-143. https://doi.org/10.1111/j.1466-8238.2009.00490.x.

668 Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their
669 probability distributions. Bull. Calcutta Math. Soc. 35, 99-109.

670 Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W.,
671 Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: An almost
672 complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic.
673 18, 31-41. https://doi.org/10.1016/j.ecolind.2011.10.009.

674 Bonada, N., Prat, N., Resh, V.H., Statzner, B., 2006. Developments in aquatic insect biomonitoring: A
675 comparative analysis of recent approaches. Annu. Rev. Entomol. 51, 495-523.
676 https://doi.org/10.1146/annurev.ento.51.110104.151124.

677 Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of southern Wisconsin.
678 Ecol. Monogr. 27, 325-349. https://doi.org/10.2307/1942268.

679 Brown, W., 1910. Some experimental results in the correlation of mental abilities. Br. J. Psychol. 3, 296-
680 322. https://doi.org/10.1111/j.2044-8295.1910.tb00207.x.

681 Bunn, S.E., Arthington, A.H., 2002. Basic principles and ecological consequences of altered flow regimes
682 for aquatic biodiversity. Environ. Manag. 30, 492-507. https://doi.org/10.1007/s00267-002-2737-0.

683 Cao, Y., Hawkins, C.P., 2011. The comparability of bioassessments: a review of conceptual and
684 methodological issues. J. N. Am. Benthol. Soc. 30, 680-701. https://doi.org/10.1899/10-067.1.

685 Caretta, M.A., Mukherji, A., Arfanuzzaman, M., Betts, R.A., Gelfan, A., Hirabayashi, Y., Lissner, T.K.,
686 Liu, J., Lopez Gunn, E., Morgan, R., Mwanga, S., Supratid, S., 2022. Water. In: Poertner, H.-O.,
687 Roberts, D.C., Tignor, M., Poloczanska, E.S., Mintenbeck, K., Alegria, A., Craig, M., Langsdorf, S.,
688 Loeschke, S., Moeller, V., Okem, A., Rama, B. (Eds.), Climate Change 2022: Impacts, Adaptation
689 and Vulnerability. Cambridge University Press, Cambridge, pp. 551-712.

690 Cha, S.-H., 2007. Comprehensive survey on distance/similarity measures between probability density
691 functions. Int. J. Math. Models Methods Appl. Sci. 1, 300-307.

692 Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., Inouye, B.D., 2011. Using null models to
693 disentangle variation in community dissimilarity from variation in alpha-diversity. Ecosphere 2,
694 art24. https://doi.org/10.1890/ES10-00117.1.

695 Cho, I.-H., Lee, Y.-W., Kim, H.-K., Kim, Y.-J., Hwang, S.-J., Won, D.-H., Noh, S.-Y., Lee, J.-K., Kim,
696 B.-H., 2019. A multimetric diatom index for biological integrity assessment of Korean streams.
697 Korean J. Environ. Biol. 37, 204-216. https://doi.org/10.11626/KJEB.2019.37.2.204.

698 Cho, I.-H., Kim, H.-K., Lee, M.-H., Kim, Y.-J., Lee, H., Kim, B.-H., 2020. The effect of monsoon rainfall
699 patterns on epilithic diatom communities in the Hantangang River, Korea. Water 12, 1471.
700 https://doi.org/10.3390/w12051471.

701 Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Leveque, C., Naiman,
702 R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity:
703 importance, threats, status and conservation challenges. Biol. Rev. 81, 163-182.
704 https://doi.org/10.1017/S1464793105006950.

705 Feio, M.J., Hughes, R.M., Callisto, M., Nichols, S.J., Odume, O.N., Quintella, B.R., Kuemmerlen, M.,
706 Aguiar, F.C., Almeida, S.F.P., Alonso-EguiaLis, P., et al., 2021. The biological assessment and
707 rehabilitation of the world's rivers: An overview. Water 13, 371. https://doi.org/10.3390/w13030371.

708 Gotelli, N.J., 2000. Null model analysis of species co-occurrence patterns. Ecology 81, 2606-2621.
709 https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2.

710 Gotelli, N.J., Graves, G.R., 1996. Null Models in Ecology. Smithsonian Institution Press, Washington,
711 DC.

712 Haase, P., Bowler, D.E., Baker, N.J., Bonada, N., Domisch, S., Garcia Marquez, J.R., Heino, J., et al.,
713 2023. The recovery of European freshwater biodiversity has come to a halt. Nature 620, 582-588.
714 https://doi.org/10.1038/s41586-023-06400-1.

715 Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson,
716 R.K., Moe, J., Pont, D., Solheim, A.L., van de Bund, W., 2010. The European Water Framework
717 Directive at the age of 10: A critical review of the achievements with recommendations for the
718 future. Sci. Total Environ. 408, 4007-4019. https://doi.org/10.1016/j.scitotenv.2010.05.031.

719 Hwang, E.-A., Cho, I.-H., Kim, H.-K., Yi, C., Kim, B.-H., 2023. The relationship between rainfall pattern
720 and epilithic diatoms in four streams of central-western Korea for three years (2013-2015). Int. J.
721 Environ. Res. Public Health 20, 4099. https://doi.org/10.3390/ijerph20054099.

722 IPCC, 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working
723 Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
724 Cambridge University Press, Cambridge.

725 Jackson, D.A., Peres-Neto, P.R., Olden, J.D., 2001. What controls who is where in freshwater fish
726 communities: The roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 58, 157-170.
727 https://doi.org/10.1139/f00-239.

728 Jun, Y.-C., Won, D.-H., Lee, S.-H., Kong, D.-S., Hwang, S.-J., 2012. A multimetric benthic
729 macroinvertebrate index for the assessment of stream biotic integrity in Korea. Int. J. Environ. Res.
730 Public Health 9, 3599-3628. https://doi.org/10.3390/ijerph9103599.

731 Karr, J.R., 1981. Assessment of biotic integrity using fish communities. Fisheries 6, 21-27.
732 https://doi.org/10.1577/1548-8446(1981)006<0021:AOBIUF>2.0.CO;2.

733 Karr, J.R., 1999. Defining and measuring river health. Freshw. Biol. 41, 221-234.
734 https://doi.org/10.1046/j.1365-2427.1999.00427.x.

735 Kelly, M.G., Whitton, B.A., 1995. The Trophic Diatom Index: A new index for monitoring eutrophication
736 in rivers. J. Appl. Phycol. 7, 433-444. https://doi.org/10.1007/BF00003802.

737 Lee, J.-H., Kim, H.G., Han, B.-H., Cho, I.-H., Kim, H.-K., Hwang, E.-A., Hwang, S.-O., Kim, B.-H.,
738 2026. Trait-environment diagnosis of ecological instability in Korean streams using benthic diatoms
739 and machine learning: a UMAP-CDI framework. Aquat. Sci. 88, 81. https://doi.org/10.1007/s00027-
740 026-01312-4.

741 Lee, S.-W., Hwang, S.-J., Lee, J.-K., Jung, D.-I., Park, Y.-J., Kim, J.-T., 2011. Overview and application
742 of the National Aquatic Ecological Monitoring Program (NAEMP) in Korea. Ann. Limnol. - Int. J.
743 Limnol. 47, S3-S14. https://doi.org/10.1051/limn/2011018.

744 Legendre, P., De Caceres, M., 2013. Beta diversity as the variance of community data: Dissimilarity
745 coefficients and partitioning. Ecol. Lett. 16, 951-963. https://doi.org/10.1111/ele.12141.

746 Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species
747 data. Oecologia 129, 271-280. https://doi.org/10.1007/s004420100716.

748 Magoulick, D.D., Kobza, R.M., 2003. The role of refugia for fishes during drought: a review and
749 synthesis. Freshw. Biol. 48, 1186-1198. https://doi.org/10.1046/j.1365-2427.2003.01089.x.

750 Manly, B.F.J., 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology, 3rd ed. Chapman
751 & Hall/CRC, Boca Raton, FL.

752 McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform manifold approximation and projection for
753 dimension reduction. arXiv 1802.03426.

754 Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B.,
755 Solymos, P., Stevens, M.H.H., Szoecs, E., et al., 2022. vegan: Community Ecology Package. R
756 package version 2.6.

757 Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg,
758 J.C., 1997. The natural flow regime. BioScience 47, 769-784. https://doi.org/10.2307/1313099.

759 Rao, C.R., 1995. A review of canonical coordinates and an alternative to correspondence analysis using
760 Hellinger distance. Qüestiió 19, 23-63.

761 Raup, D.M., Crick, R.E., 1979. Measurement of faunal similarity in paleontology. J. Paleontol. 53, 1213-
762 1227.

763 Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., Kidd, K.A.,
764 MacCormack, T.J., Olden, J.D., Ormerod, S.J., et al., 2019. Emerging threats and persistent
765 conservation challenges for freshwater biodiversity. Biol. Rev. 94, 849-873.
766 https://doi.org/10.1111/brv.12480.

767 Resh, V.H., Brown, A.V., Covich, A.P., Gurtz, M.E., Li, H.W., Minshall, G.W., Reice, S.R., Sheldon,
768 A.L., Wallace, J.B., Wissmar, R.C., 1988. The role of disturbance in stream ecology. J. N. Am.
769 Benthol. Soc. 7, 433-455. https://doi.org/10.2307/1467300.

770 Rimet, F., Bouchez, A., 2012. Life-forms, cell-sizes and ecological guilds of diatoms in European rivers.
771 Knowl. Manag. Aquat. Ecosyst. 406, 01. https://doi.org/10.1051/kmae/2012018.

772 Rosenberg, D.M., Resh, V.H., 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates.
773 Chapman & Hall, New York.

774 Simaika, J.P., Stribling, J., Lento, J., Bruder, A., Poikane, S., Moretti, M.S., Rivers-Moore, N., Meissner,
775 K., Macadam, C.R., 2024. Towards harmonized standards for freshwater biodiversity monitoring and

biological assessment using benthic macroinvertebrates. Sci. Total Environ. 918, 170360. https://doi.org/10.1016/j.scitotenv.2024.170360.

Spearman, C., 1910. Correlation calculated from faulty data. Br. J. Psychol. 3, 271-295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecol. Appl. 16, 1267-1276. https://doi.org/10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2.

Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., Tarquinio, E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. J. N. Am. Benthol. Soc. 27, 878-891. https://doi.org/10.1899/08-053.1.

Tickner, D., Opperman, J.J., Abell, R., Acreman, M., Arthington, A.H., Bunn, S.E., Cooke, S.J., et al., 2020. Bending the curve of global freshwater biodiversity loss: An emergency recovery plan. BioScience 70, 330-342. https://doi.org/10.1093/biosci/biaa002.

Watanabe, T., Asai, K., Houki, A., 1986. Numerical estimation of organic pollution of flowing waters by using the epilithic diatom assemblage - Diatom Assemblage Index (DAIpo). Sci. Total Environ. 55, 209-218. https://doi.org/10.1016/0048-9697(86)90180-4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 05SupplementaryInformation.docx
- GA.png