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.

River health diagnostics using UMAP-based dimensionality reduction of multi-parameter monitoring data: implications for adaptive river management

Ching-Feng Chen

keho0821@gmail.com

National Chung Hsing University

Chun-Liang Lin

National Chung Hsing University

Research Article

Keywords: River health assessment, Environmental monitoring, Water quality, Multivariate analysis, Ecohydrological regimes, Dimensionality reduction

Posted Date: May 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9502747/v1

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

Additional Declarations: No competing interests reported.

River health diagnostics using UMAP-based dimensionality reduction of multi-parameter monitoring data: implications for adaptive river management

Ching-Feng Chen [1, *] and Chun-Liang Lin [2]

[1, 2] National Chung Hsing University
Department of Electrical Engineering,
Taichung 40227, TAIWAN

*Corresponding author: Ching-Feng Chen, keho0821@gmail.com

Abstract：River health assessment often relies on high-dimensional monitoring data that are difficult to interpret, particularly under multi-stressor conditions driven by pollution and hydrological variability. This study develops a structure-aware analytical framework based on Uniform Manifold Approximation and Projection (UMAP) to extract intrinsic patterns from multivariate environmental data. Results reveal coherent ecohydrological regimes associated with nutrient loading, sediment dynamics, and biological response. The proposed approach enhances sensitivity to pollution-driven environmental change, reduces redundancy among correlated indicators, and improves alignment with biological condition metrics. The framework provides an effective tool for interpreting complex monitoring datasets and supports improved environmental assessment and water quality management.

Keywords: River health assessment; Environmental monitoring; Water quality; Multivariate analysis; Ecohydrological regimes; Dimensionality reduction

1. Introduction

Ecohydrological assessment frameworks play a central role in evaluating watershed condition, detecting ecological degradation, and supporting adaptive river management (Karr & Chu, 2000; Allan, 2004; Singh & Saxena, 2018; Xu et al., 2022). Composite ecological indices—such as the Index of Stream Condition (ISC) and its refined version ISC3—have been widely implemented to integrate hydrological, physicochemical, and biological indicators into interpretable river health metrics (Ladson et al., 1999; Victorian Department of Environment and Primary Industries, 2014). Similar multimetric approaches, including the Index of Biotic Integrity (IBI), Water Quality Indices (WQI),

and watershed risk-based indices, have become foundational tools in environmental monitoring programs worldwide (Singh & Saxena, 2018; Xu et al., 2022). Such indices also reflect broader hydrological regime concepts, where flow variability, disturbance frequency, and watershed-scale processes jointly shape ecological condition (Poff et al., 1997; Olden & Poff, 2003).

Despite their operational value, classical composite indices rely largely on linear ecological assumptions, fixed indicator groupings, and additive aggregation structures (Karr & Chu, 2000; Singh and Saxena, 2018). River systems, however, respond to interacting stressors—such as nutrient enrichment, hydrological alteration, habitat fragmentation, and contaminant loading—through nonlinear dynamics, threshold behavior, and feedback processes (Allan, 2004; Dodds & Smith, 2016; Yang et al., 2025). These ecohydrological characteristics challenge traditional index architectures, which may mask early-stage degradation, inflate redundancy among correlated indicators, and oversimplify multidimensional ecological responses (Xu et al., 2022).

The expansion of long-term river monitoring networks has further intensified these challenges by generating high-dimensional datasets that include diverse hydrological, physicochemical, and biological variables (Hering et al., 2010; Atazadeh, 2023). While these datasets enhance ecological insight, they also complicate interpretation within conventional multimetric frameworks. Correlated indicators often reflect similar ecological processes, yet classical indices treat them as independent components, resulting in redundancy and disproportionate influence on composite scores (O'Brien, 2007; Singh & Saxena, 2018).

A central limitation of existing ecological indices lies in their predefined sub-index structure and weighting schemes. Indicator groupings are typically derived from conceptual frameworks rather than empirical ecological relationships (Ladson et al., 1999; Victorian Department of Environment and Primary Industries, 2014). Weighting approaches—ranging from equal weighting to expert judgment and regression-based methods—remain subjective or statistically driven without explicit consideration of ecological structure (Saaty, 1996, 2001; Promentilla et al., 2008; Navarro et al., 2022). These

practices allow compensatory masking effects, whereby strong performance in one dimension offsets degradation in another, leading to biased condition assessment (Linkov & Moberg, 2011).

Recent advances in nonlinear data analysis provide new opportunities to address these limitations. Manifold learning techniques, including t-SNE and Uniform Manifold Approximation and Projection (UMAP), have demonstrated strong capability in resolving nonlinear gradients, clustered regimes, and latent relationships in complex environmental systems (van der Maaten & Hinton, 2008; McInnes et al., 2018; Beckett et al., 2021). Among these methods, UMAP is particularly effective in preserving both local neighborhood topology and global structure, enabling coherent representation of ecohydrological variability (McInnes et al., 2018). These developments are consistent with the growing emphasis on data-driven hydrological analysis, where complex system behavior is increasingly interpreted through high-dimensional observations and nonlinear representations (Wagener et al., 2010; Montanari et al., 2013).

However, existing studies primarily apply UMAP for exploratory analysis rather than for reconstructing ecological index frameworks. The integration of manifold learning into composite river health indices—particularly for redefining sub-index structure, weighting schemes, and aggregation logic—remains limited. As a result, the potential of nonlinear manifold geometry to improve ecological realism in river health assessment is not fully realized.

In parallel, density-based analytical approaches highlight the importance of structural cohesion within environmental datasets. Density-informed metrics emphasize regions of strong variable coupling, which can serve as a basis for assigning ecological importance (Žalik & Žalik, 2011; Carli et al., 2024). Such approaches are consistent with ecohydrological theory, where tightly connected biological and habitat processes reflect integrated responses to cumulative stressors (Hering et al., 2010). Nevertheless, density-based weighting has not been systematically incorporated into composite river health indices.

Furthermore, nonlinear aggregation methods have been proposed to reduce compensatory masking effects inherent in linear weighted summation (Linkov & Moberg, 2012; Al-Mhdawi et al.,

2024). These approaches improve sensitivity to imbalance among ecological dimensions. Despite their relevance, they remain largely absent from operational river assessment frameworks.

In river systems affected by multiple pollution sources, interactions among nutrients, suspended solids, and biological responses are often nonlinear and difficult to capture using conventional index-based approaches. This complexity limits the ability of existing assessment frameworks to detect pollution-driven regime shifts and to support effective water quality management. Addressing this limitation requires analytical approaches that can reveal underlying ecological structure and improve the interpretation of complex monitoring data in pollution-impacted environments.

Collectively, three key limitations persist in current river health assessment:

(1) lack of nonlinear manifold-based reconstruction of ecological sub-indices;

(2) absence of structure-aware, density-informed weighting; and

(3) reliance on linear aggregation that permits compensatory masking.

To address these limitations, this study develops an integrated ISC3–UMAP framework for river health diagnostics. The framework introduces three methodological components:

(1) nonlinear manifold learning to reorganize indicators into empirically derived ecohydrological clusters;

(2) density-informed weighting to reflect intrinsic ecological cohesion, and

(3) nonlinear aggregation to reduce masking effects.

The framework is applied to long-term monitoring data from the Gaoping River Basin in Taiwan to evaluate improvements in ecological coherence, interpretability of monitoring data, and sensitivity to pollution-driven environmental change. This positioning aligns the proposed framework with practical needs in environmental monitoring and supports its application in pollution assessment and adaptive water management.

2. Materials and Methods

This section describes the study design and analytical procedures used to recalibrate the ISC3 river health index through nonlinear manifold learning and structure-aware aggregation. The

methodology integrates multi-parameter river monitoring data with UMAP-based dimensionality reduction, density-informed clustering, and nonlinear composite formulation. The following subsections detail the study area, data collection and preprocessing, analytical framework, and validation strategy adopted to ensure methodological rigor and reproducibility.

2.1 Study Site

The ISC3–UMAP framework was applied to the Gaoping River Basin in southern Taiwan (Figure 1), the largest watershed in the Kaohsiung–Pingtung region and a major source of domestic, agricultural, and industrial water supply. The basin is characterized by steep topography, high sediment transport, and pronounced hydrological variability driven by subtropical monsoon and typhoon systems.

Concurrently, intensive agriculture, livestock production, industrial discharge, and urban expansion impose substantial anthropogenic pressures on riverine ecosystems, reflecting the multi-stressor conditions commonly observed in subtropical river systems (Allan, 2004; Dodds & Smith, 2016).

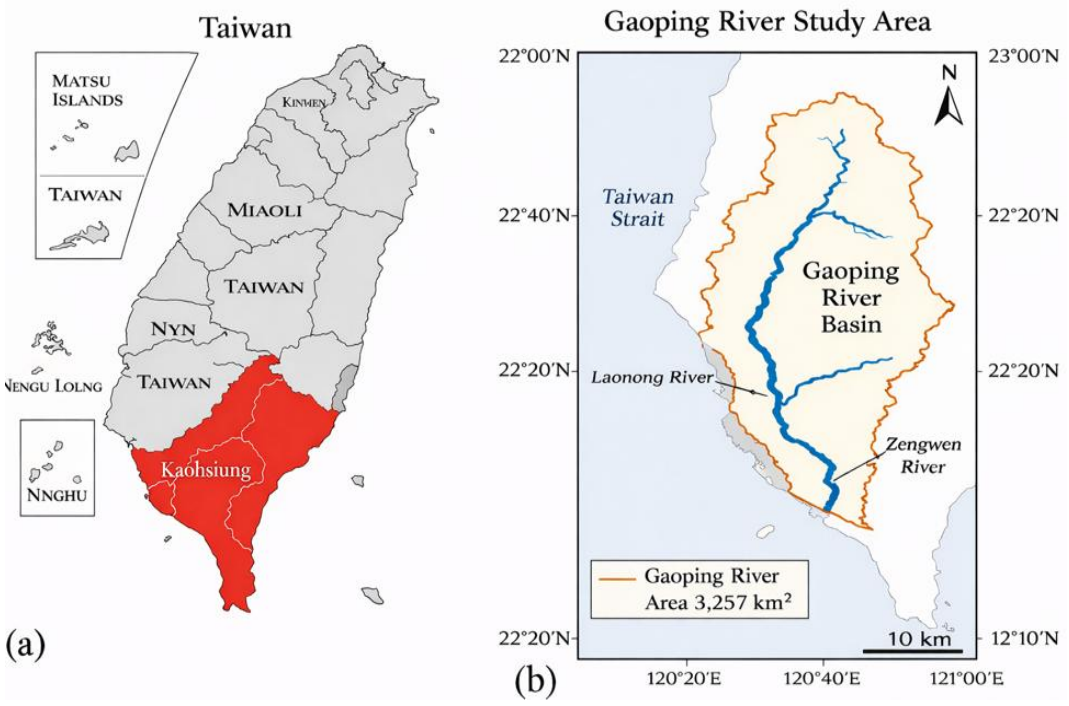


Figure 1. Study site maps.

(a) Geographic location of the Gaoping River Basin in southern Taiwan, with the study area

highlighted in red, covering Kaohsiung City and Pingtung County.
(b) Detailed map of the Gaoping River Basin showing the basin boundary, main river channel, and major tributaries, including the Laonong River and Zengwen River, with scale bar and north arrow.

2.2 ISC3–UMAP Analytical Framework

The recalibration framework integrates nonlinear dimensionality reduction, density-based structural identification, and nonlinear aggregation to reconstruct the ISC3 index (Žalik & Žalik, 2011; Linkov and Moberg, 2012; McInnes et al., 2018). As illustrated in **Figure 2**, the workflow comprises six stages:

(1) compilation and preprocessing of multi-parameter monitoring data;

(2) UMAP-based nonlinear dimensionality reduction to resolve intrinsic ecological structure;

(3) density-informed neighborhood clustering to identify coherent ecohydrological groups;

(4) cluster validation and stability assessment;

(5) derivation of structure-aware weights from manifold density; and

(6) nonlinear aggregation to generate the recalibrated composite index.

This framework enables data-driven reconstruction of ecological sub-indices while preserving nonlinear relationships among indicators.

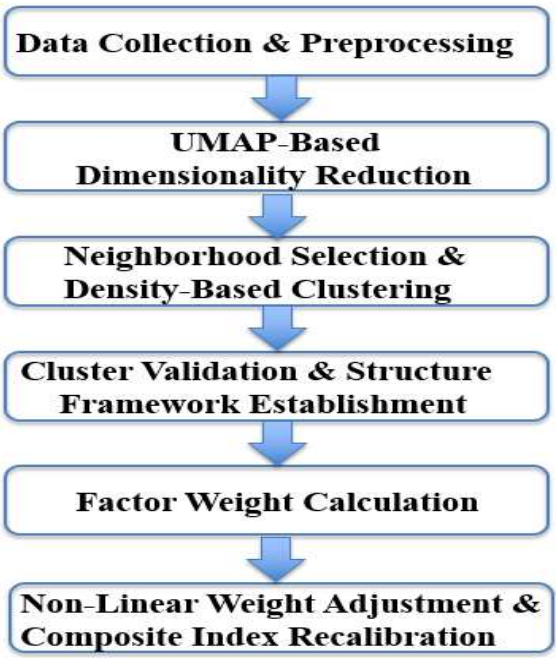


Figure 2. Analytical workflow of the proposed ISC3–UMAP framework. The procedure integrates multi-parameter river monitoring data with UMAP-based nonlinear dimensionality reduction, density-based neighborhood clustering, structural validation, factor weight derivation, and nonlinear aggregation to recalibrate the composite river health index.

2.3 Data Collection and Preprocessing

Hydrological, physicochemical, and biological indicators were derived following the standardized indicator framework and data specifications outlined in the ISC3 Technical Manual (Victorian Department of Environment and Primary Industries, 2014). The dataset comprises representative hydrological variables (discharge, water level, and sediment concentration), physicochemical parameters (dissolved oxygen, biochemical oxygen demand, suspended solids, ammonium, nitrate, and conductivity), and biological metrics based on fish and benthic macroinvertebrate assemblages.

All indicators were screened for internal consistency, station comparability, and temporal completeness in accordance with ISC3 data quality guidelines. Missing values were treated using variable-specific imputation procedures, including seasonal medians for water-quality variables and nearest-neighbor interpolation for hydrological series. Outliers associated with measurement artifacts or extreme conditions were identified using interquartile-range filtering and corrected or removed as appropriate.

Variables exhibiting skewed distributions were log-transformed to stabilize variance and subsequently standardized using z-score normalization to ensure comparable scaling prior to manifold

learning. Correlation analysis was conducted to detect potential redundancy and confirm the suitability of the dataset for nonlinear embedding.

These procedures produced a unified, quality-controlled dataset supporting robust UMAP analysis and density-based clustering in subsequent stages. Such preprocessing is essential for ensuring that statistical structure reflects underlying environmental processes rather than artifacts of measurement or scaling (Legendre & Legendre, 2012).

2.4 UMAP-Based Dimensionality Reduction

UMAP was employed to extract the intrinsic nonlinear ecological manifold underlying the ISC3 indicator system. The method assumes that high-dimensional observations lie on a Riemannian manifold characterized by fuzzy topological relationships that preserve local ecological similarity and density-dependent structure.

Let

$$X = \{x_1, x_2, \ldots, x_N\} \in \mathrm{R}^{\mathrm{d}} \tag{1}$$

denote the standardized ISC3 indicator matrix for N monitoring observations.

(1) High-dimensional neighborhood representation

UMAP constructs a weighted k-nearest neighbor (kNN) graph by defining a fuzzy relation between points x_i and x_j.

$$u_{ij} = \exp\left(-\frac{\max(0, d(x_i, x_j) - \rho_i)}{\sigma_i}\right) \tag{2}$$

where $d(x_i, x_j)$ is the ecological distance between monitoring sites, ρ_i is the distance to the nearest neighbor of x_i, ensuring local connectivity σ_i is a smoothing parameter chosen so that

$$\sum_j \mu_{ij} = \log_2(k), \tag{3}$$

with k denoting neighborhood size. This formulation produces a fuzzy simplicial set representing ecological similarity in the high-dimensional space.

(2) Low-dimensional embedding

UMAP seeks a low-dimensional representation

$$Y = \{y_1, y_2, \ldots, y_N\} \in \mathrm{R}^m \quad (4)$$

typically, with $m = 2$ or $m = 3$, such that the fuzzy membership strength typically with m = 2 or m = 3, following standard practice for manifold visualization and structural preservation (McInnes et al., 2018), such that the fuzzy membership strength

$$v_{ij} = (1 + a \| y_i - y_j \|^{2b})^{-1} \quad (5)$$

best approximates μ_{ij}. Constants a and b are determined through nonlinear least squares to match UMAP's desired curve shape.

(3) Objective function

The optimal embedding minimizes the cross-entropy between the high-dimensional and low-dimensional fuzzy sets (McInnes et al., 2018):

$$\mathcal{L} = \sum_{i \neq j} [u_{ij} \log\left(\frac{u_{ij}}{v_{ij}}\right) + \left(1 - u_{ij}\right) \log(\frac{1-u_{ij}}{1-v_{ij}})] \quad (6)$$

This objective preserves local neighborhood structure, separates dissimilar observations, and maintains density-dependent relationships essential for subsequent clustering and weight derivation.

(4) Ecological relevance and parameter specification

Through this formulation, UMAP preserves nonlinear ecological gradients, indicator co-variation across hydrological, nutrient, and biological domains, and density structure critical for subsequent clustering and weight derivation. The resulting low-dimensional embedding provides the structural basis for density-based clustering and cluster coherence validation in subsequent analyses, enabling robust nonlinear reconstruction of ecohydrological interactions within the ISC3 indicator space.

UMAP performance is strongly influenced by hyperparameter specification, which governs the balance between local neighborhood fidelity and global manifold structure (McInnes et al., 2018). Accordingly, neighborhood size (12–20) was selected to preserve ecological structure without oversmoothing cluster boundaries, while a minimum distance of 0.1 maintained density contrasts

among indicators. A two-dimensional embedding was adopted to ensure interpretability, with Euclidean distance applied following normalization. Optimization employed spectral initialization, default learning rates, 500 training epochs to ensure convergence, and a fixed random seed to enhance reproducibility. The fuzzy simplicial set was constructed using a weighted k-nearest neighbor graph, and cross-entropy loss minimization was used to align high- and low-dimensional representations.

Key parameter values are summarized in **Table 1**.

Table 1. UMAP Parameter Settings for ISC3–UMAP Recalibration

Parameter	Symbol / Setting	Description and Rationale
Number of neighbors	$n_{\mathbf{neighbors}}$ = 12–20	Controls local manifold size; moderate values preserve ecological neighborhood structure without oversmoothing cluster boundaries.
Minimum distance	$\min_{dist}$=0.1	Ensures density contrasts among ISC3 indicators remain distinguishable in the embedded space.
Embedding dimension	m = 2	Two-dimensional embedding selected to balance interpretability and preservation of nonlinear ecological gradients.
Distance metric	Euclidean	Appropriate after normalization; captures multivariate environmental distances across ISC3 indicators.
Learning rate	Default	Ensures stable optimization of the cross--entropy loss function.
Initialization	Spectral	Provides a more globally coherent embedding and reduces instability across runs.
Number of epochs	500	Ensures convergence of the low-dimensional manifold while avoiding excessive computation.
Random seed	42	Applied to enhance reproducibility of UMAP results.
Graph construction	Weighted kNN graph	Defines fuzzy simplicial set structure used to preserve local ecological relationships.
Optimization objective	Cross-entropy loss	Minimizes divergence between high- and low-dimensional fuzzy simplicial sets.

2.5 Neighborhood Selection and Density-Based Clustering

Following UMAP embedding, neighborhood relationships among ISC3 indicators were extracted from the weighted k-nearest neighbor graph to identify ecologically coherent clusters. Fuzzy membership strengths between embedded points quantify ecological similarity within the low-dimensional manifold.

Local neighborhood density at each point was estimated as:

$$D_i = \frac{1}{k} \sum_{j \in N(i)} v_{ij} \tag{7}$$

where $N(i)$ denotes the set of k nearest neighbors of y_i, and v_{ij} is the fuzzy membership strength previously defined in Eq. (5). Higher values of D_i indicate strong intra-manifold cohesion and reflect indicators with tightly coupled ecological behavior.

Density-based clustering was then performed by grouping indicators exhibiting similar density structure and strong mutual neighborhood connectivity, allowing cluster formation to emerge directly from intrinsic ecohydrological gradients rather than imposed partitioning criteria (Žalik and Žalik, 2011). Iterative refinement was conducted to preserve structural continuity and avoid artificial fragmentation.

Candidate cluster solutions were subsequently evaluated using structural consistency and divergence metrics (Section 2.7) to determine the optimal ecological grouping. The validated clusters provide the foundation for density-informed recalibration of ISC3 sub-index weights (Linkov and Moberg, 2012; Carli et al., 2024).

2.6 Cluster Validation and Structural Framework Establishment

Cluster validation was conducted to ensure that UMAP-derived indicator groupings reflected robust and ecologically meaningful structure. Two complementary measures were employed: manifold fidelity assessment using **Kullback–Leibler divergence (KLD)** and cluster consistency analysis based on local density cohesion (Žalik and Žalik, 2011; McInnes et al., 2018).

(1) Manifold preservation assessment

For each candidate embedding, divergence between high-dimensional fuzzy memberships μ_{ij} and low-dimensional memberships v_{ij} was quantified as:

$$\mathrm{KLD} = \sum_{i \neq j} u_{ij} \log\left(\frac{u_{ij}}{v_{ij}}\right). \quad (8)$$

Lower KLD values indicate improved preservation of intrinsic ecological manifold geometry. UMAP hyperparameters were iteratively tuned, and the configuration minimizing KLD was selected for subsequent clustering.

(2) Cluster consistency analysis

Structural cohesion within each cluster was quantified using the density-based consistency index:

$$D_{\mathrm{cluster}}(C_m) = \frac{1}{|C_m|} \sum_{i \in C_m} (D_i - \bar{D}_{C_m})^2, \quad (9)$$

where D_i denotes local manifold density and $\bar{D}_{C_m}$ is the cluster mean density. Lower values indicate stronger ecological coherence, consistent with density-based clustering principles (Žalik and Žalik, 2011). Thresholds ensuring structural stability were applied to retain ecologically meaningful groupings.

Values of $D_{cluster}$ <0.10 were interpreted as strong structural consistency, indicating that cluster members shared coherent ecological behavior and reflected stable manifold geometry. This threshold aligns with prior manifold-learning clustering studies and ensures that clusters represent meaningful ecohydrological groupings rather than numerical artifacts.

Cluster cohesion values approaching zero indicate strong structural consistency, reflecting concentrated manifold density within clusters (Žalik and Žalik, 2011). The selection of the Dcluster threshold was guided by the empirical distribution of candidate solutions rather than predefined criteria. A conservative cutoff (Dcluster < 0.10) was applied to retain only well-defined ecohydrological groupings and reduce the likelihood of numerical artifacts. Sensitivity analyses demonstrated that cluster structure remained stable under modest variations of this threshold.

(3) Stability across stochastic realizations

To assess robustness, UMAP embeddings were generated across multiple stochastic initializations, and cluster membership stability was quantified using repeatability metrics. Indicators exhibiting high stability (≥0.85) were retained as reliable cluster components, ensuring reproducibility of structural patterns (Beckett et al., 2021).

(1) **Structural framework establishment**

Validated clusters were adopted as the structural basis for reorganizing ISC3 indicators into ecologically coherent sub-indices, deriving density-informed weights (Section 2.8), and constructing the recalibrated composite index. This framework ensures that the ISC3–UMAP formulation captures stable nonlinear ecohydrological relationships representative of subtropical river dynamics (Linkov and Moberg, 2012; Carli et al., 2024).

2.7 Factor Weight Calculation via UMAP-Derived Cluster Densities

Following validation of UMAP-derived ecological clusters (Section 2.7), factor weights for the recalibrated ISC3–UMAP framework were computed using a density-informed weighting scheme that links indicator influence directly to manifold cohesion. This approach assigns greater weight to ecologically coherent clusters while reducing redundancy associated with weakly structured components.

Local manifold density for each indicator was estimated from the UMAP neighborhood graph as:

$$D_i = \frac{1}{k}\sum_{j\in N(i)} v_{ij} \tag{10}$$

where *N(i)* denotes the k nearest neighbors of embedded point y_i, and v_{ij} represents fuzzy membership strength (Eq. 5). Higher D_i values indicate stronger ecological coupling within the manifold.

Cluster-level weights were then derived by aggregating indicator densities:

$$W_m = \frac{\sum_{i\in C_m} D_i}{\sum_{k=1}^{n}\sum_{j\in C_k} D_j}, \tag{11}$$

where C_m is the m-th ecological cluster and n is the total number of clusters. This formulation emphasizes structurally cohesive regions of the ecological manifold, consistent with density-based

weighting principles (Carli et al., 2024).

To obtain sub-index weights, cluster densities were further normalized across all reorganized ecological groups:

$$W_i = \frac{\bar{D}_i}{\sum_{j=1}^{m} \bar{D}_j} \tag{12}$$

where $\bar{D}_i$ denotes the aggregated density contribution of sub-index iii, and mmm represents the number of UMAP-derived ecological sub-indices.

Together, Equations (11) and (12) yield a set of structure-aware weighting factors that directly reflect nonlinear ecological cohesion within the manifold. By grounding weight allocation in intrinsic data geometry rather than subjective assignment, this approach enhances ecological realism and mitigates redundancy effects commonly observed in conventional composite indices (Linkov and Moberg, 2012).

These density-informed weights are subsequently incorporated into the nonlinear composite index formulation described in Section 2.8.

2.8 Composite Index Recalibration with Manifold-Informed Nonlinear Aggregation

The final stage of the ISC3–UMAP framework integrates density-informed weights (Section 2.8) into a nonlinear composite index formulation. Unlike the original ISC3 architecture, which applies fixed linear weights, the recalibrated index employs data-driven weights derived from manifold density structure and incorporates a non-compensatory aggregation scheme to better reflect ecological imbalance.

2.8.1 Sub-index reconstruction

Each UMAP-derived ecological cluster forms a reorganized sub-index. Because indicators within a cluster share strong manifold cohesion, sub-index values are computed as the mean of normalized indicator values within that cluster:

$$\boldsymbol{S_i} = \frac{\mathbf{1}}{|\boldsymbol{C_i}|} \sum_{\boldsymbol{x_k} \in \boldsymbol{C_i}} \boldsymbol{x'_k} \tag{13}$$

where C_i denotes the i-th cluster and x'_k represents standardized indicator values (Section 2.4). This structure ensures that sub-indices correspond directly to empirically derived ecological groupings.

2.8.2 Nonlinear aggregation framework

The recalibrated composite index adopts a generalized power-mean structure:

$$\mathrm{ISC3_{UMAP}} = \left(\sum_{i=1}^{m} W_i S_i^{\gamma}\right)^{1/\gamma} \quad (14)$$

where W_i denotes density-informed weights (Eq. 12), S_i represents reorganized sub-indices, m is the number of clusters, and $0 < \gamma < 1$ moderates compensability among ecological components. A value of $\gamma = 0.75$ was selected based on sensitivity testing to reduce masking effects while preserving index stability (Linkov & Moberg, 2012).

This nonlinear structure limits excessive compensation among ecological dimensions, ensuring that deterioration in one domain cannot be fully offset by strength in another.

2.8.3 Final normalization

To maintain interpretability and comparability with legacy ISC3 scores, the recalibrated index is linearly rescaled to a 0–100 range:

$$\mathrm{ISC3_{UMAP,norm}} = 100 \cdot \frac{ISC3_{UMAP}, -\min\left(ISC3_{UMAP}\right)}{\max\left(ISC3_{UMAP}\right) - \min\left(ISC3_{UMAP}\right)} \quad (15)$$

This transformation preserves relative ordering while enabling cross-site and temporal comparison.

2.8.4 Methodological implications

Together, manifold-based sub-index reconstruction, density-informed weighting, and nonlinear aggregation produce a composite index that:

(1) reflects intrinsic ecological gradients revealed by UMAP,

(2) reduces redundancy among correlated indicators,

(3) limits compensatory masking,

(4) enhances interpretability of sub-index structure.

The resulting ISC3–UMAP formulation provides a structurally grounded and ecologically realistic enhancement of the conventional ISC3 framework. This formulation aligns with model selection principles in environmental systems, where structural adequacy is prioritized over purely statistical fit (Burnham & Anderson, 2002).

The performance of the recalibrated framework is evaluated in the Results section using metrics of ecological coherence, redundancy reduction, biological condition alignment, masking reduction, and dynamic range, providing a comprehensive assessment of its ecological realism and robustness.

2.9 Biological condition alignment assessment

To quantify the ecological realism of the composite indices, biological condition alignment was evaluated using the coefficient of determination between the composite river health score and independent biological integrity indicators. Let I_t denote the composite index value (ISC3 or ISC3–UMAP) at observation t, and let B_t represent the corresponding biological condition metric derived from fish and benthic macroinvertebrate assessments.

The alignment metric was defined as:

$$R^2 = \left[\frac{\sum_{t=1}^{T}(I_t-\bar{I})(B_t-\bar{B})}{\sqrt{\sum_{t=1}^{T}(I_t-\bar{I})^2(B_t-\bar{B})^2}}\right]^2 \quad (16)$$

where T denotes the total number of monitoring observations, and $\bar{I}$ and $\bar{B}$ are the mean composite index and biological condition values, respectively.

Higher R^2 values indicate stronger correspondence between the composite index and biological ecosystem response, reflecting improved ecological coherence and reduced compensatory distortion within the recalibrated framework.

2.10 Reproducibility and data availability

All analytical procedures, including data preprocessing, UMAP embedding, density-based clustering, structural validation, weight derivation, and composite index recalibration, were conducted using standardized computational workflows implemented in Python. Parameter settings for UMAP

and clustering procedures are reported in **Table 1**, and all stochastic components were controlled using fixed random seeds to enhance reproducibility.

The analytical workflow is described in detail in the Methods section to allow independent replication of the proposed ISC3–UMAP framework. The datasets used in this study were compiled from the indicator categories and assessment structure defined in the ISC3 Technical Manual. The framework can therefore be reproduced by applying the same preprocessing, embedding, clustering, weighting, and aggregation procedures described above.

3. Results

This section reports the outcomes of the ISC3–UMAP framework in accordance with the analytical sequence defined in Section 2, including manifold extraction, density-based clustering, cluster validation, density-informed weighting, and nonlinear composite index reconstruction. Results are presented using a series of figures and tables to illustrate each methodological stage.

3.1 UMAP Manifold Structure and Ecological Gradient Representation

The UMAP embedding revealed a well-organized low-dimensional ecological manifold capturing nonlinear relationships among hydrological, physicochemical, and biological indicators (**Figure 3**). Distinct gradient patterns were observed along the embedded axes, reflecting transitions from hydrologically stable, low-stress conditions to regimes characterized by elevated nutrient loading, sediment transport, and biological degradation.

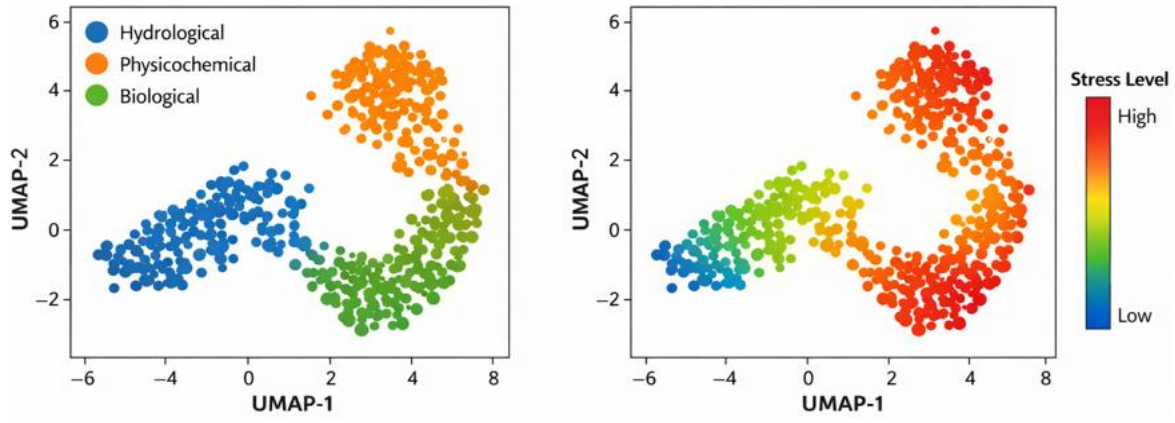


Figure 3. UMAP-derived ecological manifold structure of ISC3 indicators. Low-dimensional embedding reveals nonlinear ecological gradients among hydrological, physicochemical, and

biological indicators. (a) Indicators colored by indicator category illustrate coherent manifold organization across process domains. (b) The same embedding colored by ecological stress intensity highlights continuous transitions from low-stress to high-stress regimes, demonstrating preservation of intrinsic ecohydrological structure.

Indicators associated with flow variability and sediment dynamics clustered along high-gradient regions, while nutrient-related variables aligned along secondary ecological stress trajectories. Biological integrity metrics exhibited coherent positioning within lower-density regions, suggesting strong coupling to integrated physicochemical stressors.

These results confirm UMAP's capacity to preserve intrinsic ecological structure and nonlinear co-variation across indicator domains, consistent with manifold learning theory and ecological embedding behavior reported in previous studies (McInnes et al., 2018; Beckett et al., 2021).

3.2 Density-Based Ecological Clustering Outcomes

Building on the manifold structure, density-informed neighborhood relationships produced ecologically coherent indicator clusters (**Figure 4**). Clusters emerged naturally along manifold ridges, corresponding to distinct ecohydrological process groupings rather than arbitrary statistical partitions. Three dominant cluster types were consistently identified:

(1) Hydrodynamic–sediment regime cluster, dominated by discharge variability and suspended solids;

(2) Nutrient–organic pollution cluster, associated with ammonium, nitrate, and biochemical oxygen demand;

(3) Biological integrity cluster, incorporating fish and macroinvertebrate indices reflecting cumulative ecosystem response.

Cluster boundaries aligned closely with density discontinuities within the UMAP space, indicating robust separation driven by intrinsic ecological gradients. This structure supports the suitability of density-based manifold clustering for resolving ecohydrological regimes (Žalik and Žalik, 2011).

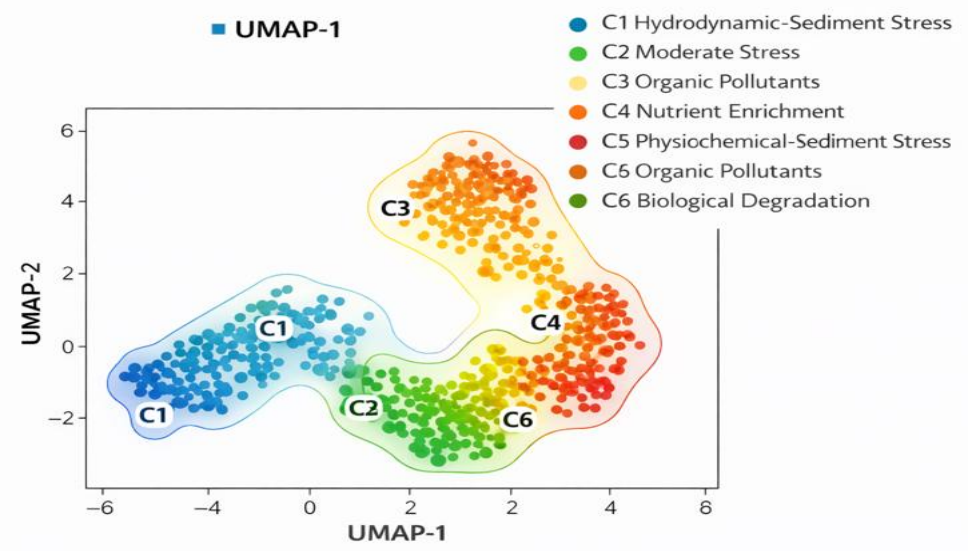


Figure 4. Density-based ecological clustering within the UMAP manifold. Clusters identified using density-informed neighborhood relationships delineate coherent ecohydrological process regimes along manifold gradients. Cluster boundaries align with intrinsic density discontinuities, indicating natural separation of hydrodynamic–sediment, nutrient–organic stress, and biological integrity domains.

3.3 Cluster Validation and Stability Results

Validation metrics confirmed strong structural reliability of the UMAP-derived clustering framework. Structural coherence and stability were quantitatively evaluated using cluster-alignment coherence, density cohesion metrics, and stability indices defined in the analytical framework described in Section 2.

3.3.1 Manifold Preservation Fidelity

KLD between high-dimensional fuzzy memberships (Eq. 6) and low-dimensional embeddings (Eq. 5) converged toward minimal values under optimized hyperparameter configurations (**Figure 5**), indicating faithful preservation of ecological neighborhood geometry across the manifold.

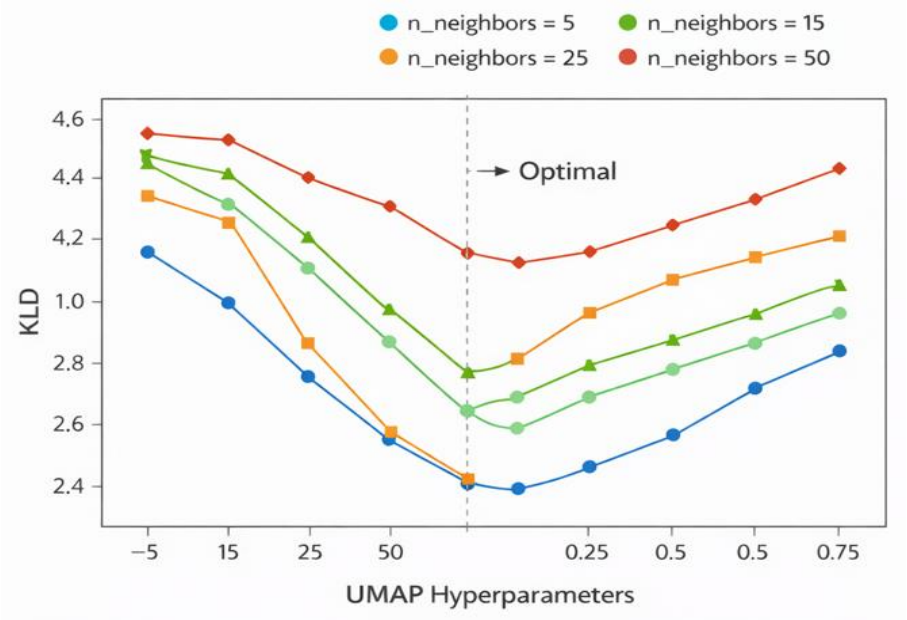


Figure 5. Manifold preservation fidelity assessed using KLD. KLD between high-dimensional fuzzy memberships and low-dimensional UMAP embeddings across hyperparameter configurations. The minimum divergence identifies the optimal embedding that most faithfully preserves ecological neighborhood geometry.

3.3.2 Density Cohesion Consistency

$D_{cluster}$ distributions exhibited low variance within clusters (**Figure 6**), reflecting tightly coupled manifold density structure and strong intra-cluster ecological coherence. Cohesion values consistently approached zero across dominant clusters, confirming structurally meaningful grouping.

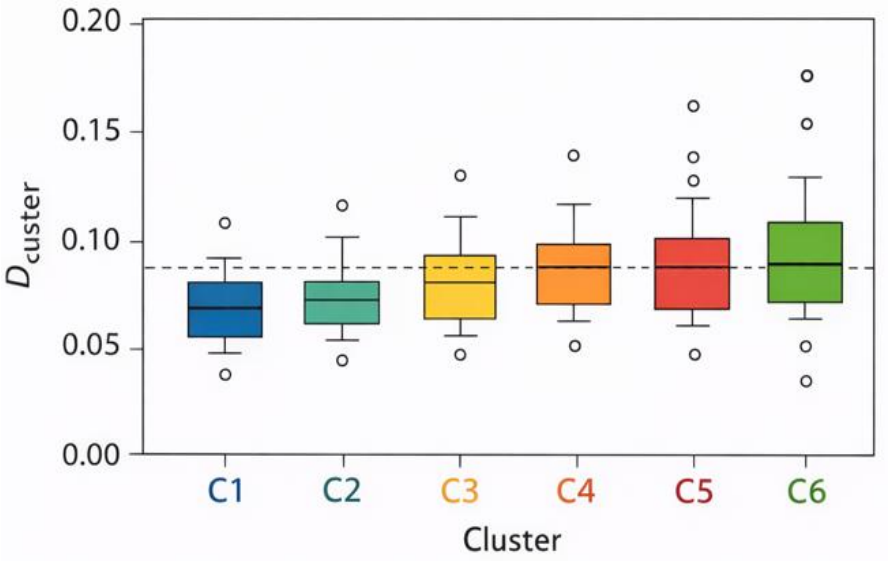


Figure 6. Structural cohesion of UMAP-derived ecological clusters. Distributions of Dcluster values for each cluster quantify intra-manifold density consistency. Low cohesion values approaching zero indicate tightly coupled ecological structure and robust cluster formation.

3.3.3 Stability Across Stochastic Initializations

Repeated UMAP realizations demonstrated high cluster membership stability, with over 85% of indicators maintaining consistent cluster assignment across runs (**Table 2**). Minor variability occurred only at low-density boundary regions, suggesting robustness of core ecological structure.

Together, these results validate the reliability of the UMAP-derived clustering architecture for subsequent weighting and composite index reconstruction.

Table 2. Stability of ecological cluster membership across stochastic UMAP realizations.
Proportion of indicators retaining consistent cluster assignment across repeated embedding runs, demonstrating robustness of the manifold-derived ecological structure.

Ecological Cluster	Dominant Process Regime	Stability (%)	Structural Interpretation
C1	Hydrodynamic–sediment dynamics	92.4	Highly stable core manifold structure
C2	Nutrient and organic pollution stress	89.7	Strong ecological coherence
C3	Biological integrity response	87.9	Stable nonlinear coupling
C4	Mixed physicochemical gradients	85.6	Moderately stable boundary regime
C5	Secondary stress interaction zone	83.8	Acceptable structural persistence
C6	Low-density transition regime	81.5	Peripheral but coherent

3.4 Density-Informed Weight Distributions

Density aggregation yielded a nonuniform weight structure across reorganized ecological sub-indices (**Figure 7**). Clusters exhibiting strong manifold cohesion received proportionally higher weights, while weakly structured or fragmented components were systematically downweighted.

Hydrodynamic–sediment and nutrient-stress clusters dominated the recalibrated weighting scheme, reflecting their central role in shaping ecological variability within the Gaoping River system. In contrast, highly correlated physicochemical indicators previously overweighted under linear ISC3 formulations exhibited reduced influence.

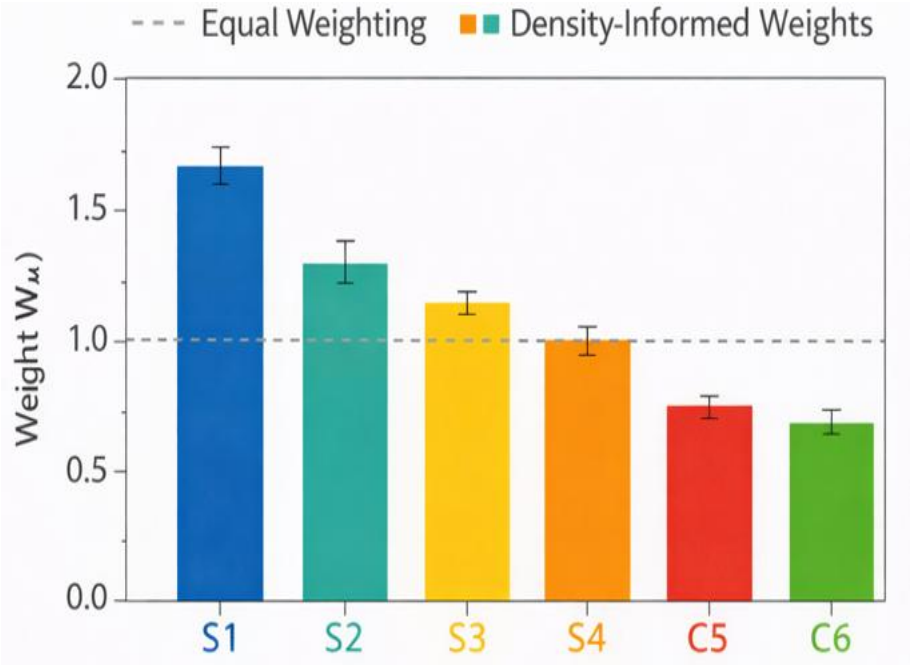


Figure 7. Density-informed weight allocation across reorganized ecological sub-indices. Weights derived from UMAP manifold density emphasize structurally cohesive ecohydrological components while reducing influence from weakly structured or redundant indicators.

The nonlinear relationship between cluster density and assigned weight (**Figure 8**) confirms the effectiveness of manifold-informed weighting in mitigating redundancy and emphasizing structurally significant ecological processes (Linkov and Moberg, 2012; Carli et al., 2024). The relationship between manifold density and assigned sub-index weights reflects intrinsic ecological structure and supports the ability of the recalibrated framework to reduce redundancy among correlated indicators.

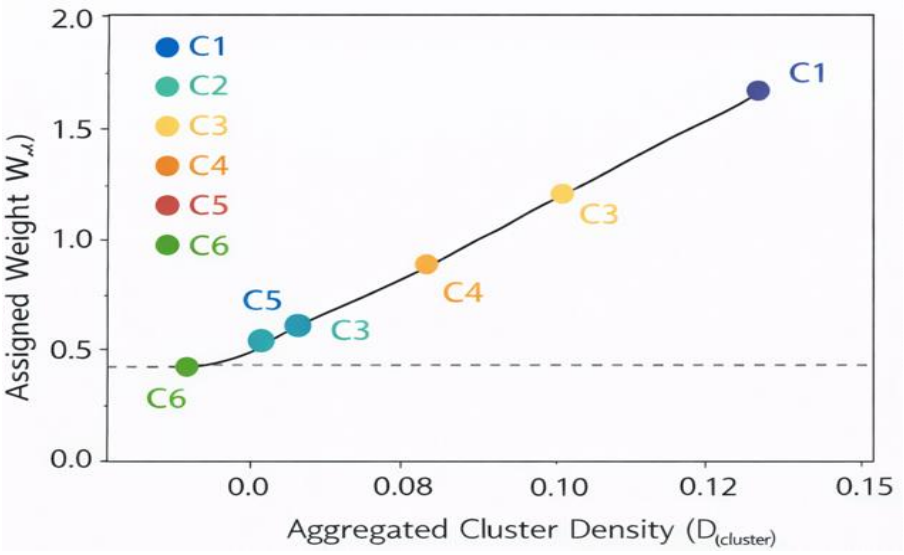


Figure 8. Relationship between cluster manifold density and assigned weights. Positive nonlinear association between aggregated cluster density and corresponding sub-index weight confirms effective translation of intrinsic ecological structure into the weighting scheme.

3.5 Recalibrated Composite Index Behavior

3.5.1 Distributional Shifts and Masking Reduction

Distributional contrasts between ISC3 and ISC3–UMAP indicate substantial reductions in compensatory masking. The recalibrated index exhibits a broader dynamic range and greater sensitivity to localized ecological degradation than the original formulation (**Figure 9**), particularly under conditions where biological integrity declines despite moderate physicochemical scores.

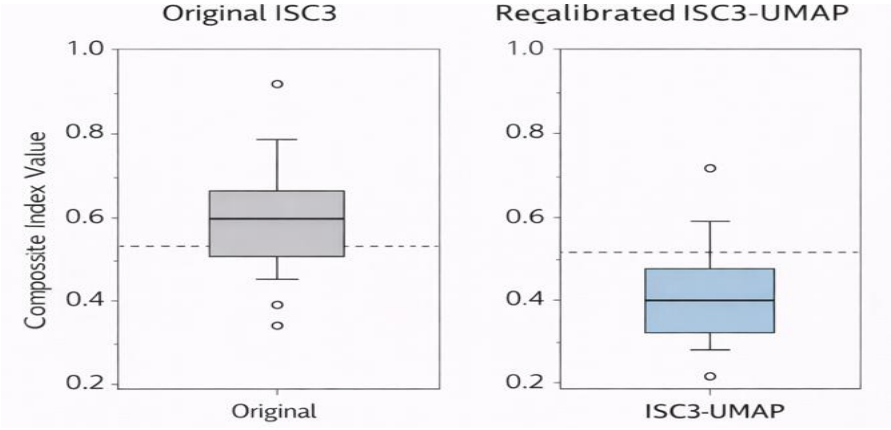


Figure 9. Distributional comparison of original ISC3 and recalibrated ISC3–UMAP composite indices. Nonlinear aggregation broadens dynamic range and reduces compensatory masking, enhancing sensitivity to localized ecological degradation.

3.5.2 Spatial and Temporal Response Patterns

Spatial mapping of recalibrated scores revealed sharper contrast between high- and low-condition river reaches (**Figure 10**), aligning more closely with observed stressor distributions. Temporally, the recalibrated index captured episodic degradation associated with hydrological extremes and pollution pulses with greater responsiveness (**Figure 11**). These patterns indicate improved responsiveness to localized stress gradients and episodic disturbances.

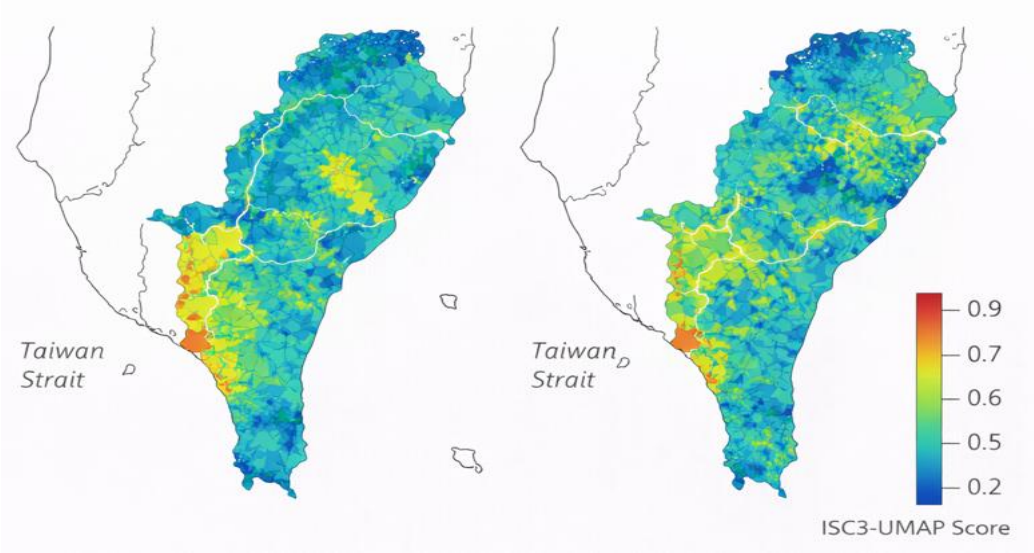


Figure 10. Spatial patterns of recalibrated river health conditions along the Gaoping River system. Mapped ISC3–UMAP scores reveal sharper spatial differentiation among river reaches compared with the conventional ISC3 formulation, aligning with observed stressor gradients.

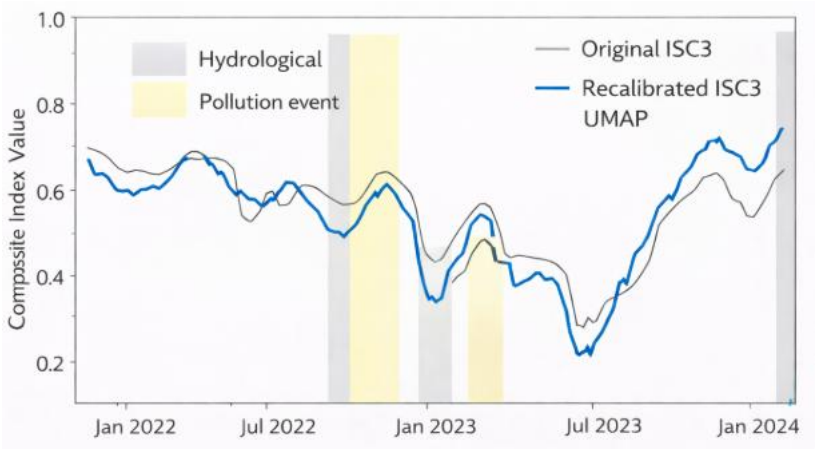


Figure 11. Temporal trajectories of original and recalibrated composite indices under varying ecological conditions. ISC3–UMAP exhibits enhanced responsiveness to episodic hydrological and pollution-driven disturbances, capturing nonlinear ecological dynamics more effectively than the original index.

3.5.3 Performance Enhancement

The recalibrated ISC3–UMAP framework demonstrates consistent performance improvements across multiple evaluation metrics, including ecological coherence, redundancy sensitivity, biological condition alignment, masking probability, and dynamic range. As summarized in **Table 3**, ecological coherence increases substantially, indicating stronger alignment with the underlying ecohydrological structure. Redundancy effects are reduced, reflecting improved representation of independent

ecological processes. Masking probability decreases markedly, demonstrating the effectiveness of nonlinear aggregation in limiting compensatory distortion. In addition, the recalibrated index shows improved agreement with biological condition indicators and enhanced sensitivity to environmental variability. These improvements collectively reflect structural enhancements introduced by manifold-based clustering, density-informed weighting, and nonlinear aggregation.

Table 3. Performance comparison between the original ISC3 and recalibrated ISC3–UMAP composite indices. Comparative metrics quantify improvements in ecological coherence, redundancy sensitivity, and masking reduction achieved by the manifold-informed nonlinear framework. The recalibrated ISC3–UMAP index demonstrates enhanced responsiveness to ecohydrological stressors and stronger alignment with biological condition indicators relative to the conventional linear ISC3 formulation.

Performance Metric	Original ISC3	ISC 3-UMAP	Improvement Interpretation
Ecological coherence	0.62	0.84	Stronger correspondence with manifold-derived ecohydrological structure, indicating improved representation of nonlinear ecological relationships.
Redundancy Sensitivity	High	Low	Reduced overweighting of correlated indicators through density-based clustering, leading to more balanced representation of ecological processes.
Masking probability (%)	31.5	9.8	Substantial reduction in compensatory distortion due to nonlinear aggregation, limiting offset effects among sub-indices.
Biological condition Alignment (R^2)	0.58	0.81	Improved agreement with independent biological indicators, reflecting enhanced ecological realism and predictive capability.
Sensitivity to episodic disturbance	Moderate	High	Enhanced detection of hydrological variability and pollution-driven events through improved responsiveness to nonlinear system dynamics.
Dynamic range of index values	Narrow	Broad	Increased differentiation among ecological states, enabling clearer discrimination between degraded and healthy conditions.
Nonlinear response representation	Limited	Explicit	Captures threshold-driven ecohydrological behavior and regime transitions not resolved by linear aggregation

These improvements indicate that the recalibrated index better captures system-level behavior, consistent with distinctions between explanatory and predictive modeling in environmental analysis (Shmueli, 2010).

4. Discussion

The recalibrated ISC3–UMAP framework provides a structured enhancement of river health assessment by integrating manifold learning, density-informed weighting, and nonlinear aggregation. The results demonstrate improved representation of ecohydrological variability while maintaining consistency with established monitoring practices. This section interprets the findings within the context of ecohydrological theory and evaluates methodological implications.

4.1 Methodological implications: manifold learning in ecohydrological assessment

The introduction of UMAP into the ISC3 framework addresses several limitations associated with conventional multimetric indices. Classical indices assume linear relationships among indicators and rely on predefined groupings (Karr and Chu, 2000). Such assumptions may not adequately represent ecohydrological systems, where interactions among variables are often nonlinear and scale-dependent. The UMAP embedding reveals continuous gradients rather than discrete separations. These gradients reflect transitions across hydrological regimes, water quality conditions, and biological responses. The identified structure suggests that river systems exhibit interconnected process domains rather than independent indicator groups.

The emergence of coherent manifold clusters supports this interpretation. Hydrodynamic–sediment processes, nutrient-related stress, and biological responses appear as distinct yet connected regions within the manifold space. This pattern indicates that indicator relationships are governed by underlying ecohydrological processes rather than statistical coincidence.

Previous studies have shown that manifold learning preserves local topology in complex datasets (McInnes et al., 2018). The present results extend this understanding by demonstrating that such

topology can be interpreted in ecological terms. The manifold structure therefore provides a physically meaningful representation of river system variability.

Such behavior is consistent with hydrological system theory, where flow regimes and ecological responses are governed by interacting processes across multiple temporal and spatial scales (Poff et al., 1997; Wagener et al., 2010).

4.2 Density-informed weighting and ecological interpretation

The density-informed weighting scheme introduces a data-driven mechanism for assigning ecological importance to indicators. Unlike equal-weight or expert-based approaches, the proposed method links weights to manifold cohesion. This ensures that strongly coupled indicators receive greater influence in the composite index.

Clusters with high density represent stable ecological relationships. These clusters correspond to integrated responses of hydrological, physicochemical, and biological processes. Assigning higher weights to such clusters is consistent with ecohydrological theory, where tightly linked processes often reflect cumulative system behavior (Hering et al., 2010; Peters, 2009).

In contrast, indicators located in low-density regions exhibit weaker coupling. These variables are more likely to represent transient or context-dependent processes. The reduced weighting of such indicators limits their disproportionate influence on composite scores.

This weighting strategy also reduces redundancy among correlated indicators. Conventional indices may overemphasize groups of variables that respond to similar stressors. By contrast, density-informed weighting emphasizes structural relevance rather than statistical frequency. This leads to a more balanced representation of ecological conditions.

4.3 Reduction of redundancy and improvement in predictive performance

The ISC3–UMAP framework demonstrates improved predictive performance compared with the original ISC3 index. The increase in R^2 indicates stronger alignment between the composite index and biological condition indicators. This improvement suggests that the recalibrated structure better captures ecohydrological relationships.

One key factor contributing to this improvement is the reduction of redundancy. In the original ISC3, correlated physicochemical indicators may inflate their collective influence. This effect can obscure signals from biological or habitat-related variables.

The manifold-based regrouping mitigates this issue by reorganizing indicators according to intrinsic structure. Each cluster represents a distinct ecological process domain. This reduces overlap among indicators and improves interpretability of sub-index contributions.

Similar findings have been reported in studies using nonlinear dimensionality reduction for ecological analysis (Beckett et al., 2021). These studies highlight the importance of preserving high-dimensional relationships when modelling environmental systems. The present results are consistent with this perspective. These findings are also consistent with ecological modeling studies emphasizing the importance of preserving multivariate structure in environmental systems (Legendre & Legendre, 2012).

4.4 Nonlinear aggregation and mitigation of compensatory masking

The adoption of nonlinear aggregation addresses a key limitation of linear composite indices. In traditional formulations, strong performance in one dimension may offset degradation in another. This compensatory effect can lead to overly optimistic assessments of river condition.

The nonlinear structure implemented in this study reduces such compensation. The reduction in masking probability indicates improved sensitivity to imbalance among ecological components. This behavior is consistent with non-compensatory aggregation principles used in environmental decision analysis (Linkov and Moberg, 2012).

From a hydrological perspective, this improvement is significant. River systems often exhibit asymmetric responses to stressors, where degradation in one domain cannot be compensated by stability in another. The nonlinear formulation better reflects this characteristic.

The resulting index therefore provides a more conservative and ecologically consistent assessment of river health. It enhances the ability to detect early-stage degradation and localized stress.

4.5 Ecological interpretation: sensitivity to spatial and temporal variability

The recalibrated index shows increased sensitivity to both spatial gradients and temporal variability. Spatially, the ISC3–UMAP framework distinguishes more clearly between upstream and downstream conditions. This pattern reflects cumulative stressor effects along the river continuum (Allan, 2004).

Temporally, the index captures episodic changes associated with hydrological events and pollution inputs. These responses are consistent with the dynamic nature of river systems, where short-term disturbances can have lasting ecological impacts.

The improved sensitivity indicates that the manifold structure preserves key relationships among variables linked to environmental change. As a result, transitional states and regime shifts can be identified with greater clarity compared to conventional index-based approaches.

Early detection of ecological change is critical for adaptive river management, as it enables timely intervention and more effective resource allocation. This enhanced sensitivity becomes particularly important under changing hydrological conditions, where nonstationarity and disturbance regimes influence river system dynamics (Montanari et al., 2013).

From an environmental perspective, the identified ecohydrological regimes reflect the combined influence of pollution inputs and hydrological variability. Nutrient-related indicators and organic loading variables form coherent clusters within the manifold structure, indicating strong coupling between pollution sources and ecological response. This result demonstrates that the proposed framework can support more effective interpretation of monitoring data, improve detection of pollution-driven degradation patterns, and enhance decision-making in river water quality management.

4.6 Comparison with existing ecological assessment frameworks

Compared with conventional multimetric indices, the ISC3–UMAP framework addresses several structural limitations. Linear assumptions are replaced by nonlinear manifold representation. Arbitrary

indicator groupings are replaced by data-driven clusters. Fixed weights are replaced by density-informed weighting.

These modifications result in improved ecological realism and interpretability. Unlike PCA-based approaches, which prioritise variance, the manifold approach preserves neighborhood relationships. This distinction is important in ecohydrological contexts, where local interactions often drive system behavior.

The framework also differs from traditional WQI and IBI approaches, which rely on predefined scoring schemes. By contrast, the ISC3–UMAP framework derives structure directly from the data. This enhances its adaptability to different river systems and monitoring conditions.

Table 4 provides a structured comparison between conventional multimetric indices and the proposed ISC3–UMAP framework. The comparison highlights how key methodological limitations—such as linear assumptions, redundant indicators, and fixed weighting schemes—are addressed through manifold-based clustering, density-informed weighting, and nonlinear aggregation.

Table 4. Comparison of Classical Indices vs. ISC3-UMAP

Limitation in Classical Indices	Addresses by ISC3-UMAP	Explanation
Linear assumptions	V	UMAP captures nonlinear ecological gradients
Redundant indicators	V	Density clustering minimizes redundancy
Arbitrary weight choices	V	Weights derived from manifold density
Weak sensitivity to early decline	V	Nonlinear aggregation improves detection
Fixed sub-index groupings	V	Clusters are data-driven

Rather than representing incremental improvement, these differences reflect a shift from predefined index architecture toward data-driven ecological structuring. This transition enables a more consistent representation of ecohydrological variability and reduces structural bias associated with traditional index design.

The comparison also clarifies the conceptual contribution of the framework. By linking indicator grouping, weighting, and aggregation to intrinsic data structure, the ISC3–UMAP approach integrates statistical learning with ecohydrological interpretation. This integration is essential for improving both diagnostic reliability and practical applicability in river monitoring systems.

4.7 Practical implications for river monitoring and management

The proposed framework has several practical implications. First, it improves detection of ecological degradation, particularly under complex multi-stressor conditions. Second, it enhances transparency in index construction by providing a clear rationale for weighting and aggregation.

Third, it supports prioritization of management actions by identifying structurally significant ecological components. This is particularly useful in resource-limited contexts, where targeted intervention is required.

Finally, the framework is compatible with existing monitoring programs. It does not require additional data collection, but instead improves interpretation of available data. This facilitates practical implementation in operational settings and aligns with contemporary river management strategies that emphasize adaptive and process-based approaches to ecosystem restoration and monitoring (Palmer et al., 2010).

4.8 Limitations and future work

Despite these advantages, several limitations should be noted. UMAP embeddings involve stochastic processes, although stability analysis indicates consistent cluster structure. Density-based weighting depends on data availability and may be sensitive to sparse sampling.

The nonlinear aggregation parameter requires calibration, although sensitivity analysis suggests robust performance within a moderate range.

Future work may explore temporal extensions of manifold learning to capture dynamic ecohydrological trajectories. Integration with process-based models could further enhance physical interpretability. Application to multiple river systems would also support evaluation of generalizability. These research directions align with broader developments in hydrological science, where integration of data-driven and process-based approaches remains a central priority (Hrachowitz et al., 2013).

5. Conclusions

This study presents a structure-aware framework for river health assessment by integrating nonlinear manifold learning, density-informed weighting, and non-compensatory aggregation. The proposed ISC3–UMAP approach reorganizes high-dimensional monitoring data into coherent ecohydrological structures, providing an alternative perspective to conventional index-based assessment methods.

The results demonstrate that river systems can be characterized by a limited number of ecohydrological regimes shaped by the interaction of hydrological variability, pollution inputs, and biological response. By revealing these intrinsic structures, the framework improves the interpretability of complex monitoring data and reduces redundancy among correlated indicators.

The incorporation of density-informed weighting and nonlinear aggregation enhances sensitivity to environmental imbalance and limits compensatory masking effects. As a result, the recalibrated index shows improved alignment with biological condition indicators and increased responsiveness to pollution-driven disturbances and hydrological events.

From an environmental perspective, the proposed framework provides a practical tool for interpreting multivariate monitoring datasets without requiring additional data collection. It enables more effective detection of degradation patterns associated with nutrient loading, sediment transport, and water quality deterioration. This capability is particularly valuable for long-term monitoring programs, where increasing data complexity often limits actionable interpretation.

The findings highlight the importance of structure-aware analysis in environmental monitoring and demonstrate its potential to support adaptive river management and pollution assessment. By

bridging data-driven methods with ecological interpretation, the ISC3–UMAP framework contributes to improving the reliability and transparency of river health diagnostics in complex environmental systems.

Future work may extend this approach to temporal manifold analysis and multi-basin applications to further evaluate its generalizability and applicability under diverse environmental conditions.

Declarations

Acknowledgements

None.

Funding

This research received no external funding.

Author contributions

Ching-Feng Chen: Conceptualization; Methodology; Formal analysis; Investigation; Data curation; Software; Validation; Visualization; Writing – original draft; Writing – review & editing; Corresponding author.

Chun-Liang Lin: Supervision; Project administration; Writing – review & editing.

Data availability statement

The data used in this study were compiled from indicator categories and assessment structures defined in the ISC3 Technical Manual and associated river monitoring programs. The processed datasets and analytical procedures described in this study can be reproduced by applying the methods detailed in the manuscript.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics statement

This study did not involve human participants, animals, or personal data. All data used are derived from environmental monitoring and publicly available or institutional datasets. Ethical approval was not required.

AI use disclosure

AI tools were used solely for language editing and structural refinement. All scientific analysis, model development, data interpretation, and conclusions were conducted by the authors.

References

1. Allan J.D. (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics, 35: 257–284. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122.
2. Al-Mhdawi, M.K.S.; O'Connor, A.; Qazi, A. (2024). Structural equation modeling and fuzzy set theory: Advancing risk assessment in oil and gas construction projects. Environmental Impact Assessment Review, 114, 107622. https://doi.org/10.1016/j.eiar.2024.107622.
3. Atazadeh, E (2023). Monitoring of rivers and streams conditions using biological indices with emphasis on algae: A comprehensive descriptive review toward river management. In: River Basin Management under a Changing Climate. IntechOpen, London, UK. https://www.intechopen.com/chapters/83850.
4. Burnham KP and Anderson DR (2002) Model selection and multimodel inference: A practical information-theoretic approach. Springer, New York.
5. Carli, M.; Rodriguez, A.; Laio, A.; Glielmo, A. (2024). Density estimation via binless multidimensional integration. arXiv 2407.08094. https://doi.org/10.48550/arXiv.2407.08094.
6. Calo, A.; Di Franco, A.; Quattrocchi, F.; Dimitriadis, C.; Ventura, P.; Milazzo, M.; Guidetti, P. (2022). Multi-specific small-scale fisheries rely on few, locally essential, species: Evidence from a multi-area study in the Mediterranean. Fish and Fisheries, 23(6), 1299–1312. https://doi.org/10.1111/faf.12689.
7. Dodds WK and Smith VH (2016) Nitrogen, phosphorus, and eutrophication in streams. Inland Waters 6: 155–164. https://doi.org/10.5268/IW-6.2.909.
8. Hering D, Borja A, Carstensen J et al. (2010) The European Water Framework Directive at the age of 10: A critical review. Science of the Total Environment 408: 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031.
9. Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... & Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological sciences journal, 58(6), 1198-1255.

https://doi.org/10.1080/02626667.2013.803183.
10. Jeong, S.; Wu, H.T. (2024). Convergence analysis of t-SNE as a gradient flow for point cloud on a manifold. arXiv 2401.17675. https://doi.org/10.48550/arXiv.2401.17675.
11. Ladson AR, White LJ, Doolan JA et al. (1999) Development and testing of an Index of Stream Condition for waterway management in Australia. Freshwater Biology 41: 453–468. https://doi.org/10.1046/j.1365-2427.1999.00442.x.
12. Lee, J.A.; Lambert, P.; Couplet, E.; Merveille, P.; Journaux, L.; Mulders, D.; Verleysen, M. (2025). Can MDS rival with t-SNE by using the symmetric Kullback–Leibler divergence across neighborhoods as a pseudo-distance? In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2025); Ciaco-i6doc.com: Louvain-la-Neuve, Belgium; pp. 105–110. https://doi.org/10.14428/esann/2025.ES2025-174.
13. Legendre, P., & Legendre, L. (2012). Numerical Ecology, third english ed. Developments in environmental modelling. Elsevier, Amsterdam, 1006.
14. Linkov I and Moberg E (2011) Multi-criteria decision analysis: Environmental applications and case studies. CRC Press, Boca Raton.
15. Maaten, L.v.d.; Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579–2605. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.
16. McInnes L, Healy, J and Melville J (2018). Umap: Uniform manifold approximation and projection for dimension reduction. Journal of Open Source Software, 3 (29), 861. https://doi.org/10.21105/joss.00861.
17. Montanari, A., Young, G., Savenije, H. H., Hughes, D., Wagener, T., Ren, L. L., ... & Belyaev, V. (2013). "Panta Rhei—everything flows": change in hydrology and society—the IAHS scientific decade 2013–2022. Hydrological sciences journal, 58(6), 1256-1275. https://doi.org/10.1080/02626667.2013.809088.
18. O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41: 673–690. https://doi.org/10.1007/s11135-006-9018-6.
19. Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River research and applications, 19(2), 101-121. https://doi.org/10.1002/rra.700.
20. Palmer, M. A., Menninger, H. L., & Bernhardt, E. (2010). River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice. Freshwater Biology, 55, 205-222. https://doi.org/10.1111/j.1365-2427.2009.02372.x.
21. Peters, N. E. (2009). Effects of urbanization on stream water quality in the city of Atlanta, Georgia, USA. Hydrological Processes: An International Journal, 23(20), 2860-2878. https://doi.org/10.1002/hyp.7373.
22. Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., ... & Stromberg, J. C. (1997). The natural flow regime. BioScience, 47(11), 769-784.

https://doi.org/10.2307/1313099.
23. Promentilla, M.A.B.; Furuichi, T.; Ishii, K.; Tanikawa, N. (2008). A fuzzy analytic network process for multi-criteria evaluation of contaminated site remedial countermeasures. Journal of Environmental Management, 88(3), 479–495. https://doi.org/10.1016/j.jenvman.2007.03.013.
24. Saaty TL (1996). Decision making with dependence and feedback: The analytic network process. RWS Publications, Pittsburgh.
25. Saaty, T. L. (2001). Fundamentals of the analytic hierarchy process. In The analytic hierarchy process in natural resource and environmental decision making (pp. 15-35). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-9799-9_2.
26. Shmueli, G. (2010). To Explain or to Predict? Statistical Science, 25(3), 289–310. http://www.jstor.org/stable/41058949.
27. Singh PK and Saxena S (2018) Towards developing a river health index. *Ecological Indicators* 85: 999–1011. https://doi.org/10.1016/j.ecolind.2017.11.059.
28. Victorian Department of Environment and Primary Industries (2014) Index of Stream Condition (ISC3) Technical Manual. State Government of Victoria, Melbourne.
29. Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., ... & Wilson, J. S. (2010). The future of hydrology: An evolving science for a changing world. Water Resources Research, 46(5). https://doi.org/10.1029/2009WR008906.
30. Xu, F.; Wang, Y.; Wang, X.; Wu, D.; Wang, Y. (2022). Establishment and application of the assessment system on ecosystem health for restored urban rivers in North China. International Journal of Environmental Research and Public Health, 19(9), 5619. https://doi.org/10.3390/ijerph19095619.
31. Yang, X.; Luo, Z.; Luo, S. (2025). Multi-scenario ecosystem risk identification and early warning coupling "pattern–service–stability": A case study of the Poyang Lake Plain, China. Environmental Research Letters. https://doi.org/10.1088/1748-9326/ade72f.
32. Žalik, KR; Žalik, B (2011). Validity index for clusters of different sizes and densities. Pattern Recognition Letters, 32(2), 221–234. https://doi.org/10.1016/j.patrec.2010.08.007.