sustainability

MDPI

Article

Impact of Environmental Factors of Stream Ecosystems on Aquatic Invertebrate Communities

Jong-Won Lee, Sang-Woo Lee, Heera Lee and Se-Rin Park *

Department of Forestry and Landscape Architecture, Konkuk University, Gwangjin-gu, Seoul 05029, Republic of Korea; jwlee8901@naver.com (J.-W.L.); swl7311@konkuk.ac.kr (S.-W.L.); heeralee@konkuk.ac.kr (H.L.)
* Correspondence: serin87@konkuk.ac.kr

Abstract: Understanding the responses of stream ecosystems to environmental disturbances is essential for maintaining and restoring healthy ecosystems. In this study, we analyzed the associations between benthic macroinvertebrate communities and environmental factors using machine learning approaches to identify key stressors potentially influencing stream ecosystem health. Various machine learning models were evaluated, with random forest (RF) and gradient boosting machine (GBM) identified as the optimal models for predicting tolerant species (TS) and Ephemeroptera, Plecoptera, and Trichoptera (EPT) species densities. SHAP analysis revealed that watershed variables, such as elevation, flow velocity, and slope, significantly influenced EPT and TS populations. EPT population density increased with elevation and flow velocity but decreased significantly with higher levels of biochemical oxygen demand (BOD), total nitrogen (TN), and agricultural land-use proportions, with negative effects becoming evident beyond threshold levels. Conversely, TS population density showed a positive response to elevated BOD, TN, and agricultural land-use proportions, stabilizing at the threshold levels of BOD and TN, but continuing to increase with greater agricultural land use. Through machine learning, this study provides critical insights into how environmental variables are associated with the distribution of benthic macroinvertebrate communities. By identifying threshold levels of key stressors, this approach offers actionable guidance for managing agricultural runoff, enhancing riparian buffers, and implementing sustainable land-use practices. These findings contribute to the development of integrated watershed management strategies that promote the long-term sustainability of stream ecosystems.

Keywords: EPT species; tolerant species; physical habitat quality; sustainable watershed management; SHAP analysis

check for updates

Academic Editor: Mirosław Grzybowski, Katja Neves and Magdalena Bowszys

Received: 19 December 2024
Revised: 16 January 2025
Accepted: 24 January 2025
Published: 4 February 2025

Citation: Lee, J.-W.; Lee, S.-W.; Lee, H.; Park, S.-R. Impact of Environmental Factors of Stream Ecosystems on Aquatic Invertebrate Communities. *Sustainability* **2025**, *17*, 1252. https://doi.org/10.3390/su17031252

1. Introduction

Stream ecosystems are complex systems in which natural processes and human activities interact and are shaped by the close relationship between watershed characteristics and stream conditions [1,2]. Human activities such as urbanization, industrialization, agriculture, and logging significantly impact stream ecosystems by altering hydrological regimes and modifying habitat structures. These changes typically lead to reduced biodiversity, increased pollution, and compromised ecosystem functions, threatening the sustainability of these vital systems. Therefore, effective stream management and restoration require integrated assessments that consider both physicochemical water quality and biological health. For example, Serpa [3] combined physicochemical, biological, and ecotoxicological approaches to assess river water quality, highlighting the importance of a comprehensive evaluation to understand the impacts of anthropogenic activities on aquatic ecosystems. Such integrated approaches are essential for promoting sustainable stream management and restoration practices that balance ecological integrity with human needs.

Understanding the associations among aquatic communities, watershed environmental changes, physicochemical variations in streams, and their responses to external stressors is critical. This has driven the development of various biological indicators for monitoring the health and biodiversity of aquatic ecosystems [4]. Periphytons, benthic macroinvertebrates, and fish are widely used in the biological assessment of aquatic ecosystems. These groups serve as ecological health indicators and offer valuable insights into the quality of aquatic habitats [4,5]. Benthic macroinvertebrates are particularly effective indicators for biological assessment. They play a key role in secondary production in stream ecosystems and occupy diverse habitats that reflect water quality and habitat structure and function [6,7]. Their community composition and structure are sensitive to varying intensities of disturbance, making them valuable tools for assessing ecological changes [8]. The advantages of using benthic macroinvertebrates as indicators include their ease of sampling, relatively short life cycles that allow consistent monitoring, and high species diversity with varying sensitivities to anthropogenic pollution [6,9]. For example, the dominance of pollution-sensitive taxa typically indicates good water quality, whereas pollution-tolerant taxa are prevalent in degraded systems [10]. Identifying these differences is crucial for assessing anthropogenic stressors and understanding causal relationships in aquatic ecosystems.

Certain benthic macroinvertebrate taxa, such as Ephemeroptera, Plecoptera, and Trichoptera (EPT), are highly sensitive to organic pollution and changes in physical habitat conditions, such as temperature fluctuations and alterations in riparian vegetation [11,12]. This sensitivity makes the EPT species particularly valuable for detecting changes in the health of stream ecosystems. In Korea, 749 EPT species have been reported, including 95 endemic species [13]. Their composition varies regionally, reflecting the differences in environmental and habitat conditions [14,15]. Conversely, taxa such as Chironomidae and Oligochaeta are more tolerant of pollution and habitat degradation, with their dominance typically signifies stream impairment. For example, saprobic values based on biochemical oxygen demand (BOD) concentrations were established for 190 taxa in Korea to classify pollution-tolerant species [16]. Comparative studies of sensitive and tolerant taxa can provide valuable insights into the specific effects of pollutants on stream ecosystems. Factors influencing benthic macroinvertebrate communities differed between the impaired and reference streams. In impaired streams, factors such as nutrient loading, sedimentation, and habitat destruction favor pollution-tolerant species that thrive under high nutrient and low oxygen conditions with elevated sedimentation [17]. In contrast, reference streams with better water quality, higher habitat complexity, and stable hydrological conditions support sensitive species such as EPT species, which require high oxygen levels and cooler, less disturbed environments [18].

Recent advances in machine learning (ML) have facilitated evaluations and predictions of the relationships between aquatic organisms and environmental factors in stream ecosystems [19–21]. ML tools are increasingly applied to model and predict the interactions among water quality, biological communities, and environmental factors, incorporating a wide range of physical, chemical, hydrological, and geomorphological variables [22]. These tools offer new insights and perspectives for river restoration and management, particularly in the context of Korean rivers, where understanding ecosystem–environment relationships remains a significant research priority [23]. Various ML models have been employed in ecological studies, including the random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost). Each model has strengths and limitations, and selecting the most suitable model for a given problem typically requires a comparative approach [20,23]. Application of multiple models allows researchers to derive optimal results while addressing complex ecological issues.

This study aimed to analyze the relationships between benthic macroinvertebrate communities and environmental factors in the Han River watershed. Specifically, we investigated the effects of environmental variables on EPT and TS, identifying the critical thresholds for these impacts. Additionally, this study compared the associations between these variables in impaired and reference streams, highlighting the differences in environmental patterns between the two stream types. We hypothesized that (1) environmental factors significantly influence the distribution of EPT and TS taxa across various stream conditions; (2) ML models effectively predict the relationships between environmental factors and benthic macroinvertebrate communities; and (3) the critical thresholds of environmental factors influencing the population densities of EPT and TS taxa differ, reflecting their contrasting ecological responses. These findings offer insights into the relationships between environmental stressors and benthic macroinvertebrate communities, providing a foundation for effective management and restoration strategies. By identifying critical ecological thresholds, this study contributes to the sustainable management of watersheds and the resilience of aquatic ecosystems.

2. Materials and Methods

2.1. Study Area

This study focuses on the upstream region of the Paldang Dam, located in the Han River Basin, which is in the center of the Korean Peninsula and is the largest watershed in South Korea (Figure 1). The upstream basin of the Paldang Dam covers an area of 23,800 km^2, and the dam, constructed in 1973, plays a vital role in supplying drinking water, generating electricity, and regulating water flow. As a key source of drinking water for Seoul and its metropolitan area, the dam is 29 m high, has a total storage capacity of 244 million tons, and provides 2.6 million tons of water per day [24]. The upstream watershed of Paldang Dam is characterized by diverse land-use types. Forests dominate the landscape, covering approximately 75.69% of the area, primarily at higher elevations. Agricultural land accounts for 16%, mainly located in the lowland regions and along riverbanks, while urban and industrial areas represent 3.1%, primarily concentrated near major roads and settlements. These land-use patterns significantly influence the water quality and ecological health of the streams, with agricultural runoff contributing to nutrient loading and urbanization, leading to increased impervious surfaces and altered hydrological flows. The streams in the study area exhibit distinct physical characteristics. Stream widths range from 6.5 to 200 m, with an average of 42.4 m, while wetted widths range from 1.5 to 118.35 m, with an averaging 12.57 m. The average water velocity is approximately 21.24 cm/s, reflecting relatively low to moderate flow rates typical of the region. The streams are generally shallow, with substrates primarily composed of gravel and sand, with occasional cobbles in higher-gradient sections and silt accumulation in lowland areas. These physical characteristics, coupled with the surrounding land use, play a critical role in shaping the benthic macroinvertebrate communities and influencing the ecological health of the watershed. Because the Paldang Dam is used as a source of drinking water, there are ongoing efforts to manage pollutants from the watershed. The upstream watershed, including the two main rivers that flow into the dam, the Bukhan and Namhan Rivers, is protected by various laws and regulations implemented by both central and local governments. These areas are designated as "water protection zones", "riparian vegetation zones", and "watershed protection zones", where land-use types and densities are regulated, and continuous efforts are made to manage water quality and reduce pollution. However, despite these efforts, large-scale construction projects, increased road density, altered hydrological patterns, and the expansion of impervious surfaces have contributed to the degradation of streams within the watershed. As a result, there is a growing need for scientifically based management strategies in the upstream region of the Paldang Dam to better address the challenges posed by watershed and climate change.

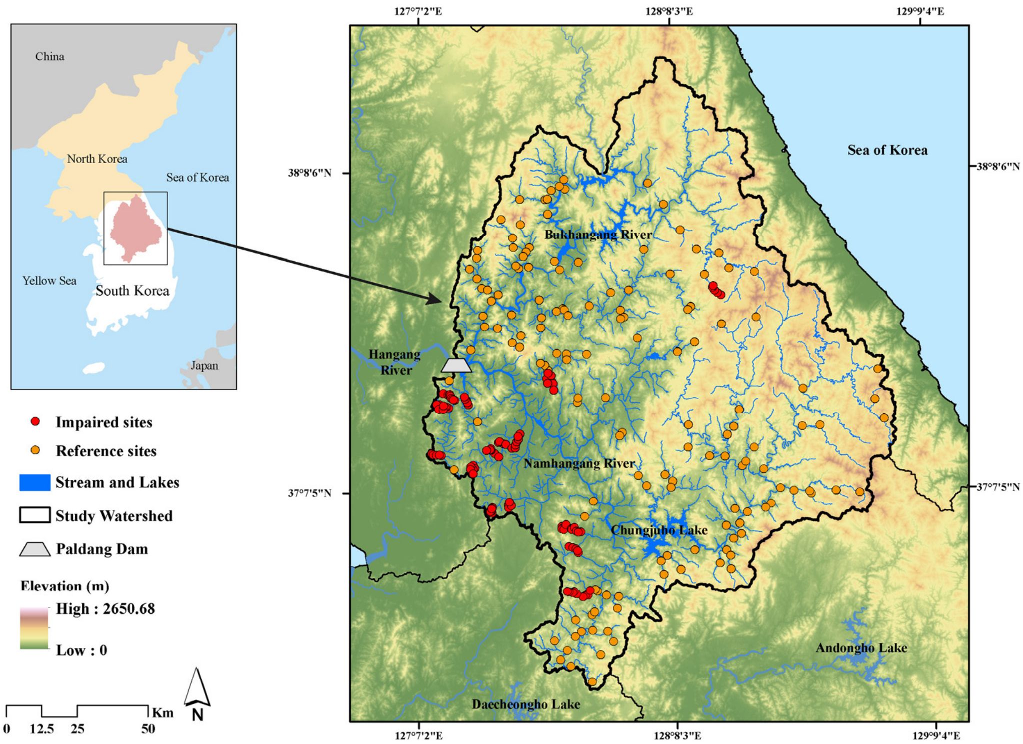

Figure 1. Location of the study areas and impaired and reference streams identified within the study areas streams.

2.2. Sampling Sites and Datasets

2.2.1. Sampling Sites

This study utilized data from 218 impaired stream sites and 148 reference stream sites collected between 2020 and 2022. The Korean Ministry of Environment comprehensively evaluates (MOE) streams and evaluates the biological health of streams through the National Aquatic Ecosystem Management Program (NAEMP) led by the National Institute of Environmental Research (NIER) [25]. Streams are classified into five grades: excellent (A), good (B), fair (C), poor (D), and very poor (E), with reference streams categorized as A or B and impaired streams as D or E [26]. Some of the impaired streams undergo field investigations through a project called "Cause Diagnosis of Stream Impairment", which aims to identify the causes of impairment prior to restoration planning. The causes of stream impairment in the study area were found to be diverse, including excessive nutrient loads from agricultural runoff, habitat destruction due to urbanization, urban non-point source pollution, and livestock-related pollution. The sampling sites were distributed across the study area at an average interval of approximately 1.8 km. The data from this project were used in this study. Reference streams are essential for comparing the extent of degradation of impaired streams and setting future goals for stream restoration and management [27]. For the analysis, we excluded streams that showed lower health from the reference stream list and used only those that met the reference criteria listed by the MOE in the Water Environment Information System. Biological, water quality, and hydrological data, excluding watershed variables, were collected in spring (April–May) and autumn (September–October).

2.2.2. Stream Datasets

Stream variables included water quality, physical habitat quality, and hydrological factors. Water quality was analyzed according to the Ministry of Environment's 'Framework Act on Environmental Policy'. In the NAEMP, water quality was evaluated by measuring the biochemical oxygen demand (BOD), ammonia nitrogen (NH_3-N), nitrate (NO_3-N), total nitrogen (TN), total phosphorus (TP), phosphate (PO_4-P), suspended solids (SS), and chlorophyll-a. This study particularly focused on variables related to organic pollutants and nutrients, which are sensitive to land use and aquatic organisms and are widely used in river management [28,29]. Additionally, dissolved oxygen (DO) and water temperature were selected because of their close relationships with organic pollution and nutrient impacts. Organic pollution is frequently associated with degraded river conditions, and BOD is a key indicator in water quality management policies in Korea [30]. The N and P concentrations in agricultural runoff play significant roles in river eutrophication and water quality degradation [31]. BOD is closely linked to DO, and temperature variations significantly influence both organic pollution and nutrient levels [32].

Physical habitat quality was assessed based on ten evaluation points: stream sandbars, naturalness, flow velocity, stream width, stream revetment, levee, substrate, cross-structure, and inland and riverside areas. However, after conducting a correlation analysis with the dependent variable (benthic macroinvertebrates), only four significant variables (stream width, substrate, riverside area, and inland) were selected for further analysis. Stream width (SW) measures the buffer zone and vegetation along the shoreline, with higher ratios indicating better habitats for aquatic plants. The stream substrate (SSE) focuses on evaluating the dominant substrate type in the streambed, such as round rocks, silt, or artificial materials such as concrete. Riverside area refers to the degree of artificiality in the dominant land use of exclusion areas, with the presence of parking lots and impermeable artificial structures indicating a high level of artificial elements. Inland represents the impact of land use within inland areas on stream ecosystems.

Water width and flow rate were selected as the hydrological variables in this study because of their substantial impact on benthic macroinvertebrate communities. These factors influence habitat availability and substrate diversity, which are key determinants of species distribution and abundance. Wider streams generally offer more diverse microhabitats, whereas the flow rate is associated with oxygenation, sediment transport, and food availability for macroinvertebrates. Variations in these variables can lead to shifts in the community structure, favoring either sensitive or more tolerant taxa depending on the conditions [33,34].

2.2.3. Watershed Datasets

Watershed variables included physiographic factors (elevation and slope) and land-use proportions. Watersheds were delineated based on sampling sites using the US EPA basins program. Using ArcGIS Pro, we calculated the average elevation and slope of each sub-watershed. The percentage of land-use types within the sub-watersheds was analyzed using the most recent high-resolution (5 m) land use and land cover (LULC) data provided by the Environmental Geographic Information Service. The LULC data were originally classified into seven categories: urban areas (including residential, industrial, commercial, transportation, and public facilities), agricultural areas (paddy fields, farms, orchards, etc.), forest areas, grassy areas, wetlands, bare soils, and water bodies. However, in this study, the sampling sites were selected based on the predominant land-use types, focusing on urban, agricultural, and forest areas.

2.2.4. Population Density Ratio of Benthic Macroinvertebrates

This study utilized data from the NAEMP and the Impairment Cause Diagnosis Project conducted by MOE and NIER. Reference stream data were obtained from the Water Environment Information System (WEIS) for surveys conducted between 2018 and 2022. Impaired stream data were derived from the results of the cause diagnosis of stream impairment project conducted during the same period [35]. Both datasets were collected in accordance with the Biological Monitoring Survey Guidelines [36].

Benthic macroinvertebrate data were collected using a Surber net with a mesh size of 1.0 mm and a frame of 3×3 cm^2. Counts were converted to density per unit area (individuals/m^2) for the quantitative assessment of survey results. Specimens were preserved in 80% ethanol solution, and species lists were prepared in accordance with the National Species List of Korea (National Institute of Biological Resources, 2018). Sampling was conducted at the most representative substrate within the survey site to ensure standardization. Whenever possible, samples were collected from riffle sections with coarse substrates; in the absence of riffles, runs and pools were sampled sequentially as alternatives. To minimize habitat disturbance, a minimum of three points were sampled, progressing upstream from downstream. Each sampling reach was approximately 100 m in length, encompassing the variability of habitats within the site. Sampling included both margins and the center of the stream channel to ensure habitat heterogeneity.

Benthic macroinvertebrate taxa belonging to the orders EPT were identified, and the EPT population density ratio was calculated. Tolerant populations were classified based on the environmental quality scores of 1 and 2 from the benthic macroinvertebrate environmental quality score list, and their densities were calculated accordingly [37]. The environmental quality score is a value that calculates the saprobic valency of each wastewater biota by distinguishing the saprobic of the sample unit and using the relative frequency of benthic macroinvertebrate taxa according to each saprobic (Supplementary Table S1).

2.3. Model Development

2.3.1. Spatial Autocorrelation Analysis

Before applying machine learning models, spatial autocorrelation analysis was conducted to assess the potential spatial patterns within the data. Moran's I was utilized to examine the global spatial autocorrelation, while local Moran's I was employed to identify localized spatial patterns in the data. The global Moran's I quantifies the degree of spatial clustering or dispersion across the dataset, whereas local Moran's I identifies significant clusters or outliers at specific locations. These analyses were conducted to ensure that the spatial structure of the data did not bias the results of the machine learning models. The spatial weights matrix was constructed using the k-nearest neighbors ($k = 5$) method, and the statistical significance of Moran's I was determined using Z-scores and p-values. Both global and local Moran's I analyses were performed using the pysal and esda Python libraries.

2.3.2. Machine Learning Models

Random forest (RF; [38]), gradient boosting machine (GBM; [39]), extreme gradient boosting (XGB; [40]), and support vector machine (SVM; [41]) models were used to investigate the relationship between EPT, TS, and environmental factors (Figure 2). The RF model is a bagging-based ensemble technique that generates bootstrap sampling using a decision-tree-based algorithm [38]. RF model has been widely applied in ecological studies involving nonlinear and complex interactions between variables [42]. The RF model creates many subsets (bootstrap sampling) from the training data and trains the same algorithm multiple times. Each predictor was estimated to determine how well it separated the two different nodes. Tree-based methods reduce trees to a size that is less likely to overfit the data, which is typically achieved through cross-validation [43]. GBM and XGB models are tree-based, similar to RF, and are ensemble techniques that sequentially combine multiple weak predictive models and weight the data incorrectly predicted to improve the error and create a strong predictive model [39,43]. The GBM model strengthens a weak model by updating the weights using gradient descent [44]. XGB is an algorithm that improves the performance of the GBM model and adds a regularization formula to the GBM model that can be overfitted. It is faster than the GBM model, suitable for complex nonlinear relationships, and has excellent parallel processing capabilities; therefore, it can solve overfitting problems that can occur in machine learning models [43]. SVM models are supervised learning algorithms based on statistical learning theory and are designed to minimize classification error rates and maximize margins [41,45]. SVMs are based on kernel functions that can solve nonlinear problems and are performed to find the decision boundary that best classifies a high-dimensional space defined by predictor variables. To determine the optimal decision boundary, a hyperplane with the maximum margin is constructed to determine the category to which a sample belongs, thereby indicating the prediction result [20,44].

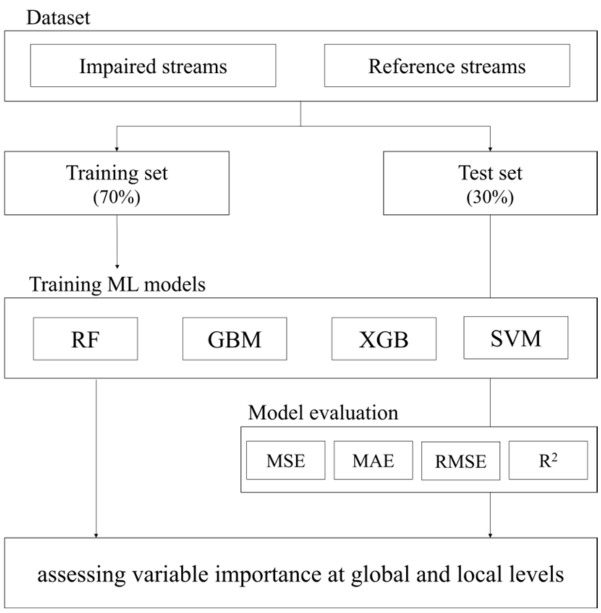

Figure 2. Modeling procedure.

Hyperparameter optimization was performed on each ML model to select the best-performing model. ML models were developed using a training set to which 70% of the samples in the dataset were randomly assigned. The remaining 30% of samples were designated as test sets. RF, GBM, and SVM algorithms were implemented using 'scikit-learn' and the XGB algorithm was implemented using 'XGBoost'; all were performed in Google Colab.

2.3.3. Model Evaluation

Appropriate evaluation metrics are important for the development of various machine learning (ML) models because they can assess the reliability and accuracy of a model. The coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean square error (MSE) are commonly used as evaluation metrics in regression models [46]. R2 is a statistical indicator that quantifies the proportion of the observed variance in the dependent variable that can be explained by the independent variables included in the model. The metric ranges from 0 to 1, with a higher value indicating greater agreement between the model and observed data. The MAE is the average of the absolute values of the differences between the predicted and observed values and serves as an indicator of the accuracy of the model when predicting variables. In contrast, RMSE calculates the square root of the average of the absolute values between the predicted and observed values. Finally, the MSE calculates the average of the absolute values of the differences between the predicted and observed values. Decreases in MAE, RMSE, and MSE indicate a higher level of model fit to the dataset. The evaluation measures selected for the ML models in this study included R^2, MAE, RMSE, and MSE [47,48].

2.3.4. SHapley Additive exPlanations (SHAP) Implementation

SHAP is an interpretable machine learning method based on the Shapley value concept of the cooperative game theory [49]. The combination of machine learning and SHAP is an effective tool for identifying important features and exploring the functional form of relationships in data [50], and has been widely applied in various fields [51]. The SHAP value is the contribution of a feature to the difference between the actual prediction and the average prediction when considering other variables in the model [52]. SHAP is useful in explaining various supervised learning models and assigns an importance value to each input variable for a specific prediction [53]. Unlike traditional features of importance in machine learning models, SHAP can identify whether the contribution of each input feature is positive or negative. In addition, each observation can be used to obtain SHAP value. Therefore, SHAP can help interpret models both locally and globally [53]. SHAP values can be approximated in various ways, such as kernel, Deep SHAP, and tree SHAP. Among these methods, Tree SHAP, a SHAP version of tree-based machine learning models (e.g., decision trees, RF, and gradient boosting trees (GBM, XGB)), was used in this study. SHAP was implemented using the SHAP 0.46.0 library.

3. Results

3.1. Descriptive Statistics

The descriptive statistics of the variables used in the analysis are listed (Table 1). The study's average BOD concentration was 2.4 mg/L, with TN and TP concentrations averaging 3.878 mg/L and 0.104 mg/L, respectively; these are classified as "slightly good" according to Korea's Environmental Policy Act. The stream width and substrate received low scores (5.17 and 13.54), indicating poor habitat conditions. The hydrological variables showed an average width of 12.9 m and a flow rate of 17.52 cm/s. Watershed analysis showed that urban and agricultural land accounted for 9.51% and 25.29%, respectively, whereas forest areas accounted for 57.90%, highlighting the significant human influence. The average EPT population density was 45.07% and the TS population density was 39.41%, suggesting relatively healthy stream ecosystems.

3.2. Spatial Autocorrelation Analysis Results

The results of the spatial autocorrelation analysis, summarized in Table 2, indicate that the global Moran's I for the residuals was calculated as −0.0327, with a Z-score of −0.284 and a *p*-value of 0.776, demonstrating that the spatial pattern of the data does not exhibit statistically significant spatial clustering or dispersion across the dataset. Similarly, the local Moran's I analysis revealed that most locations did not exhibit significant clustering or spatial outliers, with a few localized patterns (e.g., High–Low clusters) identified, though these were not statistically significant (LISA_Significance = False) (Table 3). This lack of significant spatial autocorrelation suggests that the dataset is suitable for machine learning models without requiring additional spatial adjustments, and the spatial patterns are not expected to influence model performance.

Table 1. Descriptive statistics of predictor variables.

Category	Variable (Unit)	Abbreviation	Range (Mean)	S.D.
Stream variables				
Water quality	Biochemical oxygen demand (mg/L)	BOD	0.4–10.9 (1.84)	1.36
	Total nitrogen (mg/L)	TN	0.167–12.90 (3.23)	2.15
	Total phosphorus (mg/L)	TP	0.006–0.501 (0.067)	0.082
	Dissolved oxygen (mg/L)	DO	4.74–16.31 (9.66)	2.65
	Temperature (°C)	Tem.	9.40–30.72 (18.31)	3.55
Physical habitat quality	Stream width	SW	0–25 (5.17)	3.97
	Stream substrates	SSE	0–30 (13.54)	6.74
	Riverside area	-	0–10 (7.46)	3.72
	Inland	-	0–10 (4.26)	3.29
Hydrological variables	Average water width (m)	W_Width	0.50–92.15 (12.59)	12.23
	Average velocity (cm/s)	Velocity	0.033–116.850 (17.52)	22.95
Watershed variables				
Land use	Urban area (%)	-	0.00–68.58 (9.51)	11.94
	Agricultural area (%)	-	0.00–86.07 (25.29)	19.23
	Forest area (%)	-	0.00–99.32 (57.90)	25.35
Physiographic	Average elevation (m)	-	43.30–924.20 (267.98)	198.64
	Average slope (degree)	-	1.02–20.58 (10.51)	4.72
Biological indicators				
Population density of benthic macroinvertebrates	Population density ratio of EPT species (%)	EPT	0.00–97.58 (45.07)	25.36
	Population density ratio of tolerant species (%)	TS	0.00–99.42 (39.61)	28.35

Notes: n = 366. S.D., standard deviation.

Table 2. Global Moran's I results.

Metric	Value
Moran's I	−0.0327
Expected Index	−0.0139
Variance	0.00437
Z-score	−0.284
p-Value	0.776

Table 3. Local Moran's I example results.

ID	LISA Cluster	LISA Significance	LISA Moran's I
1	High–Low	False	0.5541
13	Low–Low	False	−0.1169
20	High–Low	False	0.0420
21	High–High	False	0.0412
34	High–Low	False	0.0359

3.3. Model Structure Validation and Performance Evaluation

The performances of machine learning models, including RF, GBM, XGB, and SVM, were evaluated for EPT and TS predictions (Table 4). For EPT predictions, the RF model achieved the best performance, with the lowest MSE (254.40), RMSE (15.95), and MAE (12.74), and the highest R^2 (0.62). This highlights RF's ability to effectively model non-linear relationships and complex interactions. The GBM model followed with competitive results (MSE = 271.59, R^2 = 0.59), demonstrating strong predictive capabilities. XGB and SVM models exhibited moderate performance with higher errors (MSE = 338.04 for XGB and 307.84 for SVM) and lower R^2 values (0.49 and 0.54, respectively). For TS predictions, the GBM model outperformed the others, achieving the lowest MSE (324.48) and RMSE (18.01) and the highest R^2 (0.61). The RF model performed slightly lower but remained effective (MSE = 365.92, R^2 = 0.56). XGB and SVM models showed relatively weaker results, with MSE values of 349.89 and 429.61, respectively, and lower R^2 values. Overall, the RF model excelled in EPT predictions, while the GBM model was more effective for TS, reflecting their strengths in modeling different ecological conditions. These findings emphasize the utility of ensemble-based models like RF and GBM for robust stream ecosystem predictions, outperforming XGB and SVM in accuracy and reliability.

Table 4. Evaluation of ML model performance.

	EPT				TS			
Model	**MSE**	**RMSE**	**MAE**	**R^2**	**MSE**	**RMSE**	**MAE**	**R^2**
RF	254.40	15.95	12.74	0.62	365.92	19.13	15.19	0.56
GBM	271.59	16.48	13.02	0.59	324.48	18.01	13.88	0.61
XGB	338.04	18.39	14.95	0.49	349.89	18.71	14.85	0.58
SVM	307.84	17.55	13.53	0.54	429.61	20.73	16.15	0.48

3.4. Variable Importance at Global and Local Levels

In the SHAP summary plot, each point represents a SHAP value that measures the contribution of an environmental variable to an individual prediction at a site, whereas a cluster of points represents the magnitude, commonality, and direction of the global effect of the variable. The average SHAP value across all sites represented a measure of global importance (Figure 3). The predictors are listed from top to bottom in descending order of global importance.

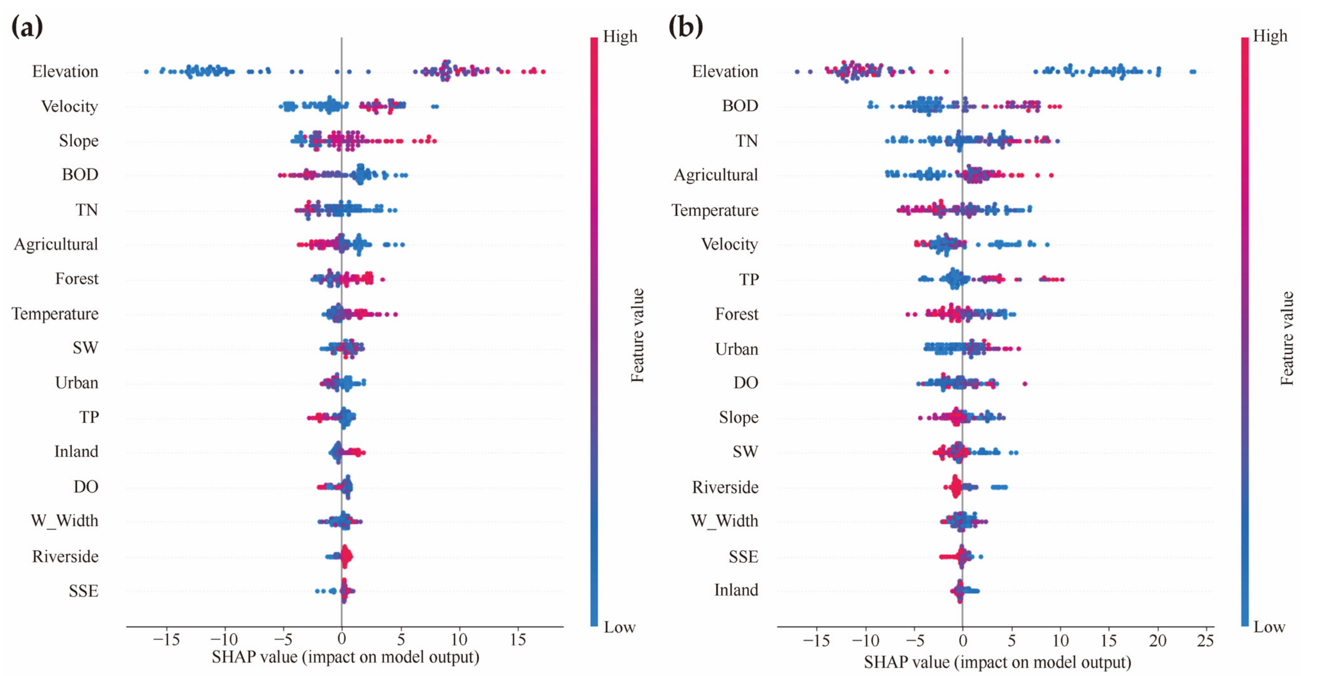

Figure 3. SHAP summary plot for (**a**) EPT (RF model) and (**b**) TS (GBM model).

The top five variables influencing the EPT population density ratio were elevation, velocity, slope, BOD, and TN, while for the TS population density ratio, the most influential variables were elevation, BOD, TN, agricultural land use, and temperature. Common variables for both the EPT and TS included elevation, BOD, and TN. The EPT population density increased with elevation but decreased with higher BOD and TN concentrations, whereas TS showed the opposite trend. Velocity and slope also positively affected EPT population density, whereas the agricultural ratio and temperature had greater effects on TS population density. Habitat variables were less significant.

The SHAP force plots for high EPT in impaired streams, EPT in reference streams, and TS population density in the dataset are shown in Figure 4. Values in bold indicate the predictions generated during model training. These values describe how individual environmental variables contribute to EPT and TS population densities and interact with each other to influence the results at a particular data point. Overall, watershed variables such as elevation, flow velocity, and slope had the greatest impact on both EPT and TS population densities. Lower elevations negatively affected EPT population density, whereas higher elevations had a negative impact on TS population density. In the reference streams, the agricultural ratio and physical habitat quality (SW and SSE) positively influenced the EPT population density, whereas flow velocity was a key positive factor in the impaired streams. In contrast, TS population density was positively influenced by the urban ratio and water quality (BOD). In impaired streams, land use and physical habitat quality (SW, inland) played a significant role in increasing TS population density. These plots provide a regional description of individual samples, illustrate the varying effects of variables on predicting EPT and TS population densities, and reveal the complex relationships between environmental factors and benthic macroinvertebrates.

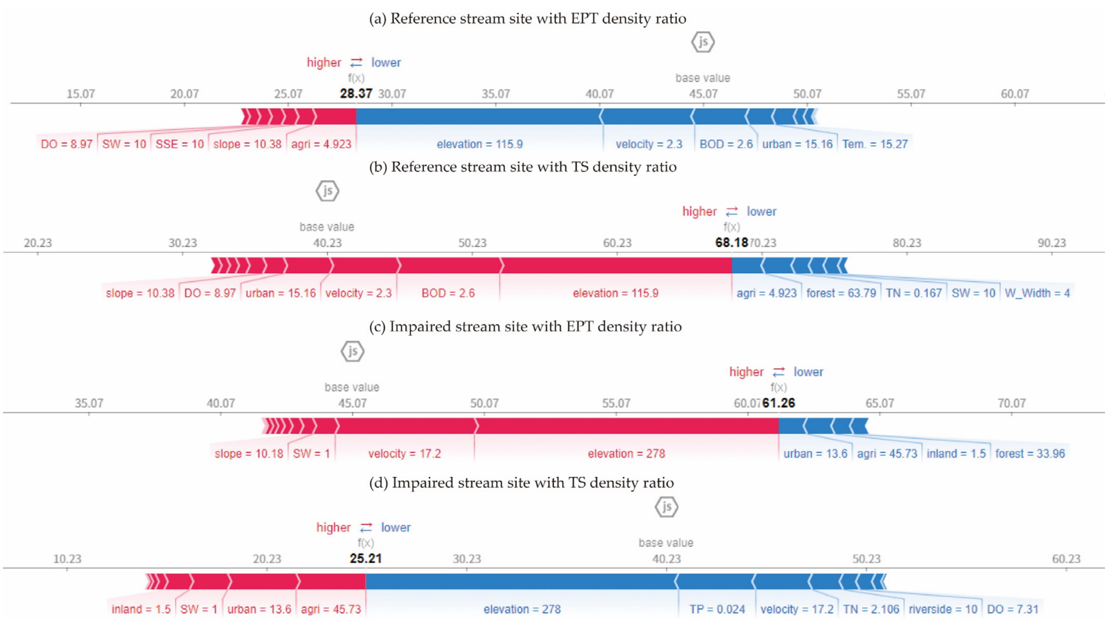

Figure 4. Force plots on a particular example: (**a**,**b**) EPT and TS population densities in the same reference stream, (**c**,**d**) EPT and TS population densities in the same impaired stream. The magnitude and direction of the SHAP values for the predictor variables are indicated by arrows (red arrows indicate a positive effect, blue arrows indicate a negative effect) to indicate their contribution (black bold values) to the final prediction.

3.5. Interpretation of Partial Effects of Environmental Parameters

The threshold values of the predictors that could change the density ratios of EPT and TS were identified using SHAP dependence plots (Figure 5). For the EPT population density ratio, negative SHAP values were observed when BOD concentrations were approximately 1.6 mg/L, TN concentrations were approximately 3 mg/L, and the proportion of agricultural area exceeded 35%. The EPT population density ratio stabilized when the BOD was approximately 3 mg/L, TN was approximately 5 mg/L, and the agricultural area proportion was approximately 40%. For the TS population density ratio, positive SHAP values were observed when BOD was approximately 2 mg/L, TN was approximately 2 mg/L, and the proportion of agricultural area exceeded 15%. The TS population density ratio stabilized at a BOD concentration of 5 mg/L and TN concentration of 6 mg/L, with the TS population density continuing to increase as the proportion of agricultural area increased.

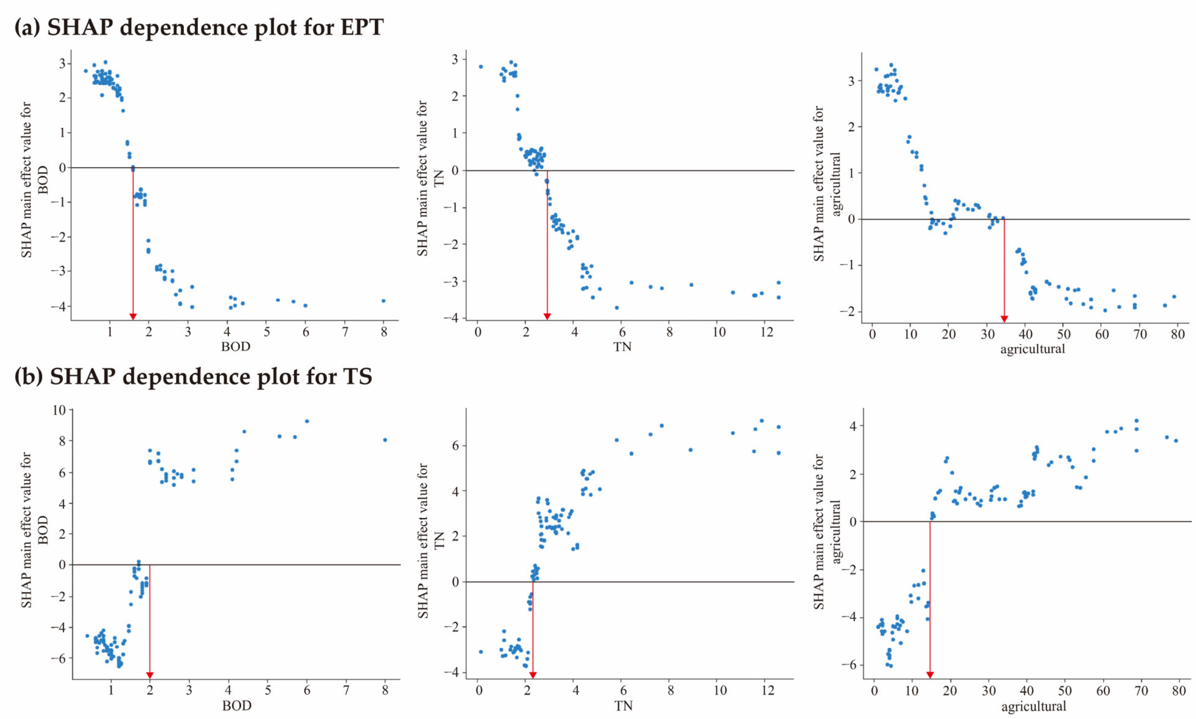

Figure 5. SHAP dependence plot for EPT. Each data point in the scatter plot represents an example of the dataset. In the SHAP summary plot, the top five significant variables are shown on the x-axis, and the SHAP values associated with those variables are shown on the y-axis. The plot shows the contribution of the feature to the model prediction as the value increases. Positive SHAP values indicate that the feature contributes to increasing the prediction, while negative SHAP values indicate that the feature reduces the prediction. Red arrows indicate threshold points in the SHAP value, highlighting critical values where the impact on the prediction shifts significantly.

4. Discussion

This study found that benthic macroinvertebrate communities are affected by environmental factors such as water quality, habitat, geomorphology, and land use. The two major biological groups of benthic macroinvertebrates (EPT and TS) responded differently to these factors.

4.1. Relationship Between Physiographic Factors and Benthic Macroinvertebrate Communities in Stream Ecosystems

This study found that altitude was a significant factor affecting EPT and TS. Altitude is significantly associated with the distribution of aquatic organisms living in streams [54]. This is consistent with previous studies showing that EPT abundance is higher in forest streams, which are generally characterized by high elevations [55]. Elevation is related to slope and geology and is inversely related to retention time, with fast-flowing watersheds tending to be higher and steeper. In particular, elevation is a major determinant of other environmental factors, such as its influence on climatic variables, including water temperature, sunlight, and air pressure, and on land use and stream conditions through its influence on geology, soil, and vegetation [56]. Slope is a key environmental variable influencing the distribution of benthic macroinvertebrates through its effects on hydrology, sediment dynamics, and physical habitat characteristics. Steeper slopes are generally associated with higher flow velocities, which create fast-flowing habitats that favor taxa adapted to such conditions, such as certain EPT species. These species often rely on coarse substrates, such as cobbles and gravels, that are more prevalent in high-gradient streams where fine sediment deposition is minimized [57]. In contrast, lower slopes are often characterized by slower flow velocities, increased sedimentation, and finer substrates, which can lead to habitat conditions favoring tolerant taxa, such as Chironomidae and Oligochaeta. These taxa are better adapted to environments with higher organic matter deposition and reduced oxygen levels, which are often associated with degraded stream conditions [58]. Moreover, slopes indirectly affect macroinvertebrate distributions by influencing riparian vegetation and land use. Steeper slopes are less likely to be developed for agriculture or urbanization, often maintaining more intact riparian buffers that contribute to improved water quality and habitat structure [59]. Conversely, flatter slopes are more susceptible to anthropogenic impacts, such as sediment and nutrient runoff, which further shape the composition of benthic communities [12]. These ecological interpretations help explain the observed relationships between slope and macroinvertebrate distributions in this study, emphasizing the importance of slope as both a direct and indirect determinant of stream health.

The correlation analysis supported these findings, showing significant positive correlations between EPT population density and elevation ($r = 0.62$) and slope ($r = 0.54$). Conversely, TS population density was negatively correlated with elevation ($r = -0.50$) and slope ($r = -0.45$) (Supplementary Table S2). These results align with the observed ecological patterns, where higher elevations and steeper slopes support conditions favorable for EPT species while limiting TS taxa.

4.2. Relationship Between Water Quality and Benthic Macroinvertebrate Communities in Stream Ecosystems

BOD and TN exhibited strong associations with EPT and TS, similar to previous studies [55,60], demonstrating how nutrients and organic pollution directly affect benthic macroinvertebrate communities. BOD serves as a key indicator of organic pollution by measuring the amount of dissolved oxygen required by aerobic organisms to break down organic matter in water. Elevated BOD levels are typically associated with increased organic matter, typically from sources such as sewage discharge, agricultural runoff, and industrial waste [61]. As BOD increases, microbial activity intensifies, consuming dissolved oxygen (DO) and potentially leading to oxygen depletion or even an anoxic state, significantly affecting aquatic life, including benthic macroinvertebrates [62]. However, in flowing water systems, the relationship between BOD and DO is influenced by additional factors. Turbulence and reaeration in streams can enhance oxygen diffusion from the atmosphere, partially offsetting the effects of increased microbial oxygen consumption [63]. Similarly, flow velocity can play a critical role in maintaining DO levels by facilitating mixing and reducing stagnant zones where oxygen depletion is more likely [64]. Consequently, while elevated BOD poses a risk of DO depletion, its impact may vary depending on the hydrodynamic characteristics of the stream. Such conditions favor TS, such as Chironomidae and Tubificidae, whereas sensitive taxa, such as EPT, are typically excluded. The influence of BOD on the ecosystem may vary depending on other factors, such as nutrient concentrations. In streams with high TN levels, nutrients interact with organic pollutants to exacerbate DO depletion. For example, increased nutrient loads can accelerate eutrophication, leading to algal blooms that deplete oxygen as they decompose, further reducing the DO levels [65]. High levels of TN were particularly influential, consistent with previous research showing that increased nitrogen concentrations are associated with a reduction in sensitive EPT species and a corresponding increase in tolerant species such as Tubificidae and Chironomidae [66]. This indicated that nutrient enrichment and organic pollution play pivotal roles in shaping benthic macroinvertebrate communities. BOD, although important, may act in conjunction with other factors, such as TN and flow dynamics, to influence the overall ecological health of streams. In this study, although BOD was strongly associated with both EPT and TS, the relative importance of DO was low. This may be explained by the fact that while DO is critical for the survival of aquatic organisms and plays a fundamental role in shaping the structure and function of benthic communities [67], its influence can be mitigated by other environmental factors. For example, higher flow velocities in certain stream sections can promote oxygenation by increasing turbulence and aeration, which may buffer the negative effects of high BOD levels [68]. Therefore, while BOD is a critical measure of organic pollution and is strongly associated with changes in benthic macroinvertebrate communities, its role must be considered alongside other variables, such as flow velocity and nutrient levels, which can influence DO and overall stream conditions. These combined factors contribute to the complex dynamics governing the distribution of EPT and TS taxa in freshwater ecosystems.

The importance of these elevations supports the findings of previous studies that highlight their important role in shaping macroinvertebrate distribution [69]. EPT generally depends on low water temperatures and high oxygen levels in streams at high elevations and gradients [69]. In contrast, TS, such as Chironomidae and Oligochaeta, are typically better adapted to a wider range of environmental conditions but may be reduced in streams with high gradients and velocities [70]. Therefore, the observed importance of slope and elevation in shaping the community structure supports the understanding that these factors are critical for maintaining habitat conditions suitable for EPT. In these environments, the EPT is better adapted to clinging to or sheltering within the substrate, minimizing the risk of displacement by the current [71]. In contrast, tolerant species such as Chironomidae are typically more abundant in low-velocity streams, where fine sediments and organic material accumulate, providing ample resources for these species [70].

Water temperature is an environmental factor associated with the metabolic rates of aquatic organisms, particularly macroinvertebrates [72]. It is a critical determinant of the distribution, survival, and reproductive success of these organisms [72]. However, the results of this study show a different pattern from those of previous studies [73]. The density of pollution-tolerant species decreased when the water temperature exceeded approximately 18 °C, contrary to the expectation that these species would thrive under warmer conditions. Piggott et al. [74] found that water temperature had various effects on benthic macroinvertebrates and that the overall richness of benthic macroinvertebrates decreased with increasing temperature, but species abundance showed a positive trend. However, the local response to temperature changes can vary significantly depending on additional factors, including dissolved oxygen and land-use patterns, as suggested by Verberk et al. [75]. This indicates that other factors, such as dissolved oxygen levels, habitat structure, or even localized land use, may interact with temperature to influence species distribution.

Benthic macroinvertebrates are sensitive to habitat conditions and are commonly used as indicators of environmental changes [76]. However, in the present study, habitat variables were relatively less important in explaining the distribution of EPT and TS. Typically, habitat factors such as substratum, riverside land, and stream width directly influence the richness and abundance of benthic invertebrates [11,77] and are typically reported to have a greater impact than other environmental variables [78]. For example, Rico et al. [79] found that chemical water quality variables had a relatively lower influence on benthic invertebrate community changes than habitat characteristics in the Danube River. The correlation analysis revealed that BOD was negatively correlated with EPT population density ($r = -0.62$) and positively correlated with TS population density ($r = 0.73$). TN also showed similar trends, with a negative correlation with EPT population density ($r = -0.46$) and a positive correlation with TS population density ($r = 0.65$) (Supplementary Table S2). These findings highlight the detrimental effects of organic pollution and nutrient enrichment on sensitive taxa such as EPT, while favoring tolerant taxa such as Chironomidae. However, some studies have reported that water quality explains the variation in benthic invertebrate communities better than habitat variables such as physical conditions and sediment characteristics [80]. These findings suggest that although habitat variables are important environmental factors for benthic invertebrates, attention must be paid to the combined effects of different variables. Habitat characteristics and physicochemical parameters simultaneously influence the health of benthic invertebrate communities, and the complexity of ecosystems can be reflected in the combined effects of these factors on their structural and functional aspects [81].

4.3. Environmental Influences on EPT and TS in Impaired and Reference Streams

In stream ecosystems, a comparison of impaired and reference streams revealed significant differences in the influence of environmental factors on the distribution of EPT species and TS. This highlights the crucial roles of local environmental conditions, land use, and nutrient levels in shaping macroinvertebrate communities in different stream environments.

In the reference streams, the EPT and TS were strongly influenced by physical characteristics, such as high elevation, steep slopes, and fast flow velocity. These streams typically experience minimal human disturbance, high forest cover, low agricultural activity, and limited urban impact [82]. These factors create an ideal habitat for EPT, which relies on well-oxygenated velocity-flowing water and complex substrate structures for survival and reproduction. In contrast, TS was less abundant in these pristine conditions as it tended to thrive in more nutrient-rich, slow-flowing waters typically associated with human-altered landscapes [83]. In impaired streams, especially those at low elevations, agriculture-dominated areas, land use, and nutrient enrichment were the dominant factors influencing the benthic macroinvertebrate community structure. Agricultural runoff, particularly nitrogen, significantly increased TS abundance owing to elevated nutrient levels and altered flow conditions, which created a more favorable environment for tolerant species.

In the reference streams, relatively undisturbed land use, including high forest cover and limited urban and agricultural activities, led to healthier macroinvertebrate communities dominated by EPT species. These streams typically have minimal nutrient inputs, leading to low TN concentrations and supporting the EPT populations. However, in agricultural regions located at higher elevations, such as high-altitude farming areas, the intensification of land use contributes to elevated nutrient levels, particularly nitrogen, which increases the abundance. In impaired streams, land-use patterns across the watershed, riverside land, and inland areas had a significant impact on benthic macroinvertebrate communities. Low agricultural and riverside land use and inland land-use intensity were correlated with decreased EPT population density. In contrast, impaired streams in agricultural regions exhibit higher nutrient loads and lower habitat complexities, resulting in increased TS abundance. This suggests that the EPT is positively affected by undisturbed high-altitude and steep-slope stream environments; however, its distribution is negatively affected by changes in watershed land use, such as agriculture or urban development.

4.4. Limitations of Current Indicators and Future Directions

While EPT and TS indicators provide valuable insights into the health of stream ecosystems, they may not fully capture the complexity and functional diversity of benthic macroinvertebrate communities. However, these indicators may overlook key aspects of ecosystem function, such as energy flow, nutrient cycling, and functional diversity, which are critical for understanding ecosystem resilience and processes. In future research, the inclusion of complementary ecological metrics, such as functional community analysis, diversity indices, or trait-based approaches, could enhance the assessment of ecosystem health. Functional community analysis, for example, evaluates the roles and interactions of species within ecosystems, providing insights into their ecological roles and resilience to environmental changes. Such integrative approaches could offer a broader perspective, helping to better capture the multifaceted impacts of environmental stressors on stream ecosystems.

5. Conclusions

In conclusion, the distribution of EPT and TS was strongly associated with key environmental factors, land use, and nutrient dynamics in both reference and degraded streams. Due to their sensitivity, EPT serves as an early warning indicator of ecological stress, while TS reflects prolonged exposure to pollution or habitat degradation. Based on the findings of this study, restoration projects aimed at increasing EPT populations should prioritize reducing nutrient inputs; maintaining TN levels below 5 mg/L is recommended to support the recovery of EPT populations. Enhancing habitat complexity through reforestation or the establishment of riparian buffer zones is also essential for providing the structural and hydrological conditions required by EPT species. In addition, for managing TS populations and mitigating the impacts of nutrient enrichment, controlling agricultural runoff is crucial. The results suggest that TS densities increased significantly when the proportion of agricultural land use exceeded 15% and stabilized at approximately 40%. Targeting these thresholds through sustainable agricultural practices, such as cover cropping, precision fertilization, and the creation of riparian buffers, can effectively reduce nutrient loading and improve stream health. These findings provide actionable guidance for watershed management, emphasizing the importance of maintaining nutrient levels within ecological thresholds and adopting land-use practices that promote sustainable and resilient stream ecosystems.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su17031252/s1. Table S1: Environmental quality score of benthic macroinvertebrates, including information on tolerant taxa; Table S2: Full correlation matrix of environmental variables and macroinvertebrate metrics.

Author Contributions: Conceptualization, J.-W.L.; Methodology, J.-W.L.; Data collection, J.-W.L.; Formal analysis, J.-W.L.; Visualization, J.-W.L.; Writing—Original draft preparation, J.-W.L.; Writing—Review and editing, S.-W.L., S.-R.P., H.L. and J.-W.L. All authors have read and agreed to the published version of the manuscript.

Funding: The Korea Environment Industry and Technology Institute (KEITI) through The Decision Support System Development Project for Environmental Impact Assessment funded by the Korea Ministry of Environment (MOE) (No. 2020002990009).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data will be made available upon request.

Acknowledgments: This research was supported by the Korea Forest Service (Korea Forestry Promotion Institute), under project number FTIS 2021331A00-2223-AA01; the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2022R1I1A1A01070904).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TS	Tolerant Species
EPT	Ephemeroptera, Plecoptera, and Trichoptera
RF	Random Forest
SHAP	SHapley Additive exPlanations
MOE	The Korean Ministry of Environment
NAEMP	The streams through the National Aquatic Ecosystem Management Program
NIER	The National Institute of Environmental Research

References

1. Allan, J.D. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **2004**, *35*, 257–284. [CrossRef]
2. Carpenter, S.R.; Stanley, E.H.; Vander Zanden, M.J. State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes. *Annu. Rev. Environ. Resour.* **2011**, *36*, 75–99. [CrossRef]
3. Serpa, D.; Keizer, J.J.; Cassidy, J.; Cuco, A.; Silva, V.; Gonçalves, F.; Cerqueira, M.; Abrantes, N. Assessment of River Water Quality Using an Integrated Physicochemical, Biological and Ecotoxicological Approach. *Environ. Sci. Process. Impacts* **2014**, *16*, 1434–1444. [CrossRef] [PubMed]
4. Birk, S.; Bonne, W.; Borja, A.; Brucet, S.; Courrat, A.; Poikane, S.; Solimini, A.; Van De Bund, W.; Zampoukas, N.; Hering, D. Three Hundred Ways to Assess Europe's Surface Waters: An Almost Complete Overview of Biological Methods to Implement the Water Framework Directive. *Ecol. Indic.* **2012**, *18*, 31–41. [CrossRef]
5. Park, S.R.; Kim, S.; Lee, S.W. Evaluating the Relationships between Riparian Land Cover Characteristics and Biological Integrity of Streams Using Random Forest Algorithms. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3182. [CrossRef] [PubMed]
6. Simaika, J.P.; Stribling, J.; Lento, J.; Bruder, A.; Poikane, S.; Moretti, M.S.; Rivers-Moore, N.; Meissner, K.; Macadam, C.R. Towards Harmonized Standards for Freshwater Biodiversity Monitoring and Biological Assessment Using Benthic Macroinvertebrates. *Sci. Total Environ.* **2024**, *918*, 170360. [CrossRef] [PubMed]
7. Johnson, R.C.; Jin, H.; Carreiro, M.M.; Jack, J.D. Macroinvertebrate Community Structure, Secondary Production and Trophic-level Dynamics in Urban Streams Affected by Non-point-source Pollution. *Freshw. Biol.* **2013**, *58*, 843–857. [CrossRef]
8. Williams-Subiza, E.A.; Brand, C.; Miserendino, M.L. Compositional Shifts in Freshwater Macroinvertebrate Communities over 30 Years of Urbanization. *Ecol. Eng.* **2022**, *183*, 106738. [CrossRef]
9. Iwasaki, Y.; Schmidt, T.S.; Clements, W.H. Quantifying Differences in Responses of Aquatic Insects to Trace Metal Exposure in Field Studies and Short-Term Stream Mesocosm Experiments. *Environ. Sci. Technol.* **2018**, *52*, 4378–4384. [CrossRef]
10. Keke, U.N.; Omoigberale, M.O.; Ezenwa, I.; Yusuf, A.; Biose, E.; Nweke, N.; Edegbene, A.O.; Arimoro, F.O. Macroinvertebrate Communities and Physicochemical Characteristics along an Anthropogenic Stress Gradient in a Southern Nigeria Stream: Implications for Ecological Restoration. *Environ. Sustain. Indic.* **2021**, *12*, 100157. [CrossRef]
11. Fierro, P.; Bertrán, C.; Tapia, J.; Hauenstein, E.; Peña-Cortés, F.; Vergara, C.; Cerna, C.; Vargas-Chacoff, L. Effects of Local Land-Use on Riparian Vegetation, Water Quality, and the Functional Organization of Macroinvertebrate Assemblages. *Sci. Total Environ.* **2017**, *609*, 724–734. [CrossRef]
12. Friberg, N. Impacts and Indicators of Change in Lotic Ecosystems. *Wiley Interdiscip. Rev. Water* **2014**, *1*, 513–531. [CrossRef]
13. National Geography Information Institute (NGII). *The National Atlas of Korea II*; Ministry of Land Infrastructure and Transport: Sejong, Republic of Korea, 2020.
14. Peredo Arce, A.; Kail, J.; Tasser, E.; Feio, M.J.; Palt, M.; Schletterer, M. The Effect of Riparian Forest on Landscape Connectivity for the EPT Community across European Regions. *Hydrobiologia* **2023**, 1–18. [CrossRef]
15. Vadas, R.L., Jr.; Hughes, R.M.; Bae, Y.J.; Baek, M.J.; Gonzáles, O.C.B.; Callisto, M.; de Carvalho, D.R.; Chen, K.; Ferreira, M.T.; Fierro, P. Assemblage-Based Biomonitoring of Freshwater Ecosystem Health via Multimetric Indices: A Critical Review and Suggestions for Improving Their Applicability. *Water Biol. Secur.* **2022**, *1*, 100054. [CrossRef]
16. Kong, D.; Son, S.-H.; Hwang, S.-J.; Won, D.H.; Kim, M.C.; Park, J.H.; Jeon, T.S.; Lee, J.E.; Kim, J.H.; Kim, J.S. Development of Benthic Macroinvertebrates Index (BMI) for Biological Assessment on Stream Environment. *J. Korean Soc. Water Environ.* **2018**, *34*, 183–201. [CrossRef]
17. Lagauzère, S.; Pischedda, L.; Cuny, P.; Gilbert, F.; Stora, G.; Bonzom, J.-M. Influence of *Chironomus riparius* (Diptera, Chironomidae) and *Tubifex tubifex* (Annelida, Oligochaeta) on Oxygen Uptake by Sediments. Consequences of Uranium Contamination. *Environ. Pollut.* **2009**, *157*, 1234–1242. [CrossRef]
18. Poulton, B.C.; Tao, J. Evaluation of EPT Macroinvertebrate Metrics in Small Streams Located within the Non-Connected Stormwater Management Region of Kansas City, Missouri, USA. *Trans. Mo. Acad. Sci.* **2019**, *47*, 21–34. [CrossRef]

19. Giri, S.; Zhang, Z.; Krasnuk, D.; Lathrop, R.G. Evaluating the Impact of Land Uses on Stream Integrity Using Machine Learning Algorithms. *Sci. Total Environ.* **2019**, *696*, 133858. [CrossRef] [PubMed]
20. Cha, Y.; Shin, J.; Go, B.; Lee, D.-S.; Kim, Y.; Kim, T.; Park, Y.-S. An Interpretable Machine Learning Method for Supporting Ecosystem Management: Application to Species Distribution Models of Freshwater Macroinvertebrates. *J. Environ. Manag.* **2021**, *291*, 112719. [CrossRef] [PubMed]
21. Lee, J.-W.; Park, S.-R.; Lee, S.-W. Effect of Land Use on Stream Water Quality and Biological Conditions in Multi-Scale Watersheds. *Water* **2023**, *15*, 4210. [CrossRef]
22. Meng, W.; Zhang, N.; Zhang, Y.; Zheng, B. Integrated Assessment of River Health Based on Water Quality, Aquatic Life and Physical Habitat. *J. Environ. Sci.* **2009**, *21*, 1017–1027. [CrossRef]
23. Yu, Q.; Ji, W.; Prihodko, L.; Ross, C.W.; Anchang, J.Y.; Hanan, N.P. Study Becomes Insight: Ecological Learning from Machine Learning. *Methods Ecol. Evol.* **2021**, *12*, 2117–2128. [CrossRef] [PubMed]
24. Ministry of Environment (MOE). *Evaluation of the Implementation of the Han River Middle Zone Water Environment Management Plan*; Han River Basin Environmental Office: Gyeonggi-do, Republic of Korea, 2022.
25. Lee, S.-W.; Hwang, S.-J.; Lee, J.-K.; Jung, D.-I.; Park, Y.-J.; Kim, J.-T. Overview and Application of the National Aquatic Ecological Monitoring Program (NAEMP) in Korea. *Ann. Limnol. J. Limnol.* **2011**, *47*, S3–S14. [CrossRef]
26. Ministry of Environment (MOE); National Institute of Environmental Research (NIER). *SOP, Standard Operation Procedure*; NIER-GP: Incheon, Republic of Korea, 2023.
27. Lee, J.-W.; Lee, S.-W.; Hwang, S.-J.; Jang, M.-H.; Won, D.-H.; An, K.-J.; Park, H.-J.; Lee, J. Establishing Diagnosis Systems for Impaired Stream Ecosystem Using Stream/River Ecosystem Survey and Health Assessment. *Korean J. Ecol. Environ.* **2020**, *53*, 1–10. [CrossRef]
28. Gerth, W.J.; Li, J.; Giannico, G.R. Agricultural Land Use and Macroinvertebrate Assemblages in Lowland Temporary Streams of the Willamette Valley, Oregon, USA. *Agric. Ecosyst. Environ.* **2017**, *236*, 154–165. [CrossRef]
29. Mouri, G.; Takizawa, S.; Oki, T. Spatial and Temporal Variation in Nutrient Parameters in Stream Water in a Rural-Urban Catchment, Shikoku, Japan: Effects of Land Cover and Human Impact. *J. Environ. Manag.* **2011**, *92*, 1837–1848. [CrossRef] [PubMed]
30. Jun, Y.-C.; Kim, N.-Y.; Kwon, S.-J.; Han, S.-C.; Hwang, I.-C.; Park, J.-H.; Won, D.-H.; Byun, M.-S.; Kong, H.-Y.; Lee, J.-E. Effects of Land Use on Benthic Macroinvertebrate Communities: Comparison of Two Mountain Streams in Korea. *Ann. Limnol. J. Limnol.* **2011**, *47*, S35–S49. [CrossRef]
31. Herringshaw, C.J.; Stewart, T.W.; Thompson, J.R.; Anderson, P.F. Land Use, Dtream Habitat and Benthicinvertebrate Assemblages in a Highly Sltered Iowa Watershed. *Am. Midl. Nat.* **2011**, *165*, 274–293. [CrossRef]
32. de Anda, J.; Gradilla-Hernández, M.S.; Díaz-Torres, O.; Díaz-Vázquez, D. Seasonal and Long-Term Behavior of TN: TP Ratio in Lake Cajititlán and Its Environmental Implications. *Water Air Soil Pollut.* **2022**, *233*, 99. [CrossRef]
33. Blöcher, J.R.; Ward, M.R.; Matthaei, C.D.; Piggott, J.J. Multiple Stressors and Stream Macroinvertebrate Community Dynamics: Interactions between Fine Sediment Grain Size and Flow Velocity. *Sci. Total Environ.* **2020**, *717*, 137070. [CrossRef]
34. Md Rawi, C.S.; Al-Shami, S.A.; Madrus, M.R.; Ahmad, A.H. Biological and Ecological Diversity of Aquatic Macroinvertebrates in Response to Hydrological and Physicochemical Parameters in Tropical Forest Streams of Gunung Tebu, Malaysia: Implications for Ecohydrological Assessment. *Ecohydrology* **2014**, *7*, 496–507. [CrossRef]
35. Ministry of Environment (MOE). *Stream Aquatic Ecosystem Health Impaired Diagnosis Guideline*; MOE: Sejong, Republic of Korea, 2019.
36. National Institute of Environmental Research (NIER). *Guidelines on Aquatic Ecosystem Status Survey and Health Assessment Methods*; National Institute of Environmental Research (NIER): Sejong, Republic of Korea, 2024.
37. Kong, D.; Park, Y.; Jeon, Y. Revision of Ecological Score of Benthic Macroinvertebrates Community in Korea. *J. Korean Soc. Water Environ.* **2018**, *34*, 251–269. [CrossRef]
38. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
39. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
40. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the KDD'16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
41. Cortes, C. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
42. Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random Forests for Classification in Ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef] [PubMed]
43. Sahin, E.K. Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. *SN Appl. Sci.* **2020**, *2*, 1308. [CrossRef]
44. Cui, Y.; Cai, M.; Stanley, H.E. Comparative Analysis and Classification of Cassette Exons and Constitutive Exons. *BioMed Res. Int.* **2017**, *2017*, 7323508. [CrossRef]

45. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; ISBN 1475732643.
46. Harishkumar, K.S.; Yogesh, K.M.; Gad, I. Forecasting Air Pollution Particulate Matter (PM2. 5) Using Machine Learning Regression Models. *Procedia Comput. Sci.* **2020**, *171*, 2057–2066.
47. Hair, J.F.; Ringle, C.M.; Sarstedt, M. Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Plan.* **2013**, *46*, 1–12. [CrossRef]
48. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Trans. ASABE* **2015**, *58*, 1763–1785.
49. Lundberg, S. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874.
50. Ij, H. Statistics versus Machine Learning. *Nat. Methods* **2018**, *15*, 233.
51. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable Ai: A Review of Machine Learning Interpretability Methods. *Entropy* **2020**, *23*, 18. [CrossRef]
52. Lundberg, S.M.; Lee, S.-I. Consistent Feature Attribution for Tree Ensembles. *arXiv* **2017**, arXiv:1706.06060.
53. Mangalathu, S.; Hwang, S.-H.; Jeon, J.-S. Failure Mode and Effects Analysis of RC Members Based on Machine-Learning-Based SHapley Additive ExPlanations (SHAP) Approach. *Eng. Struct.* **2020**, *219*, 110927. [CrossRef]
54. ABDO, A.S.; Md Rawi, C.S.; Ahmad, A.H.; Rosmahanie Madrus, M. Biodiversity of Stream Insects in the M Alaysian P Eninsula: Spatial Patterns and Environmental Constraints. *Ecol. Entomol.* **2013**, *38*, 238–249. [CrossRef]
55. Chun, S.-P.; Jun, Y.-C.; Kim, H.-G.; Lee, W.-K.; Kim, M.-C.; Chun, S.-H.; Jung, S.-E. Analysis and Prediction of the Spatial Distribution of EPT (Ephemeroptera, Plecoptera, and Trichoptera) Assemblages in the Han River Watershed in Korea. *J. Asia-Pac. Entomol.* **2017**, *20*, 613–625. [CrossRef]
56. Arocena, R.; Chalar, G.; Pacheco, J.P. Agriculture and Elevation Are the Main Factors for Pampasic Stream Habitat and Water Quality. *Environ. Monit. Assess.* **2018**, *190*, 245. [CrossRef] [PubMed]
57. Mishra, A.S.; Nautiyal, P. Longitudinal Distribution of Benthic Macroinvertebrate Assemblages in a Central Highlands River, the Tons (Central India). *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **2013**, *83*, 47–51. [CrossRef]
58. Leigh, C.; Datry, T. Drying as a Primary Hydrological Determinant of Biodiversity in River Systems: A Broad-scale Analysis. *Ecography* **2017**, *40*, 487–499. [CrossRef]
59. Feld, C.K.; Fernandes, M.R.; Ferreira, M.T.; Hering, D.; Ormerod, S.J.; Venohr, M.; Gutiérrez-Cánovas, C. Evaluating Riparian Solutions to Multiple Stressor Problems in River Ecosystems—A Conceptual Study. *Water Res.* **2018**, *139*, 381–394. [CrossRef] [PubMed]
60. Hu, X.; Zuo, D.; Xu, Z.; Huang, Z.; Liu, B.; Han, Y.; Bi, Y. Response of Macroinvertebrate Community to Water Quality Factors and Aquatic Ecosystem Health Assessment in a Typical River in Beijing, China. *Environ. Res.* **2022**, *212*, 113474. [CrossRef] [PubMed]
61. Mamun, M.; An, K.-G. Stream Health Assessment Using Chemical and Biological Multi-Metric Models and Their Relationships with Fish Trophic and Tolerance Indicators. *Ecol. Indic.* **2020**, *111*, 106055. [CrossRef]
62. Dauer, D.M.; Weisberg, S.B.; Ranasinghe, J.A. Relationships between Benthic Community Condition, Water Quality, Sediment Quality, Nutrient Loads, and Land Use Patterns in Chesapeake Bay. *Estuaries* **2000**, *23*, 80–96. [CrossRef]
63. Gibson, C.A.; Meyer, J.L.; Poff, N.L.; Hay, L.E.; Georgakakos, A. Flow Regime Alterations under Changing Climate in Two River Basins: Implications for Freshwater Ecosystems. *River Res. Appl.* **2005**, *21*, 849–864. [CrossRef]
64. Allan, J.D.; Castillo, M.M. *Stream Ecology: Structure and Function of Running Waters*; Springer: Dordrecht, The Netherlands, 2007.
65. Wang, X.; Tan, X. Macroinvertebrate Community in Relation to Water Quality and Riparian Land Use in a Substropical Mountain Stream, China. *Environ. Sci. Pollut. Res.* **2017**, *24*, 14682–14689. [CrossRef] [PubMed]
66. Su, P.; Wang, X.; Lin, Q.; Peng, J.; Song, J.; Fu, J.; Wang, S.; Cheng, D.; Bai, H.; Li, Q. Variability in Macroinvertebrate Community Structure and Its Response to Ecological Factors of the Weihe River Basin, China. *Ecol. Eng.* **2019**, *140*, 105595. [CrossRef]
67. Karaouzas, I.; Theodoropoulos, C.; Vourka, A.; Gritzalis, K.; Skoulikidis, N.T. Stream Invertebrate Communities Are Primarily Shaped by Hydrological Factors and Ultimately Fine-Tuned by Local Habitat Conditions. *Sci. Total Environ.* **2019**, *665*, 290–299. [CrossRef]
68. Álvarez-Cabria, M.; Barquín, J.; Juanes, J.A. Microdistribution Patterns of Macroinvertebrate Communities Upstream and Downstream of Organic Effluents. *Water Res.* **2011**, *45*, 1501–1511. [CrossRef]
69. He, F.; Wu, N.; Dong, X.; Tang, T.; Domisch, S.; Cai, Q.; Jähnig, S.C. Elevation, Aspect, and Local Environment Jointly Determine Diatom and Macroinvertebrate Diversity in the Cangshan Mountain, Southwest China. *Ecol. Indic.* **2020**, *108*, 105618. [CrossRef]
70. Elbrecht, V.; Beermann, A.J.; Goessler, G.; Neumann, J.; Tollrian, R.; Wagner, R.; Wlecklik, A.; Piggott, J.J.; Matthaei, C.D.; Leese, F. Multiple-stressor Effects on Stream Invertebrates: A Mesocosm Experiment Manipulating Nutrients, Fine Sediment and Flow Velocity. *Freshw. Biol.* **2016**, *61*, 362–375. [CrossRef]
71. Vimos-Lojano, D.J.; Martínez-Capel, F.; Hampel, H. Riparian and Microhabitat Factors Determine the Structure of the EPT Community in Andean Headwater Rivers of Ecuador. *Ecohydrology* **2017**, *10*, e1894. [CrossRef]

72. Bonacina, L.; Fasano, F.; Mezzanotte, V.; Fornaroli, R. Effects of Water Temperature on Freshwater Macroinvertebrates: A Systematic Review. *Biol. Rev.* **2023**, *98*, 191–221. [CrossRef]
73. Pan, B.-Z.; Wang, Z.-Y.; Li, Z.-W.; Lu, Y.-J.; Yang, W.-J.; Li, Y.-P. Macroinvertebrate Assemblages in Relation to Environments in the West River, with Implications for Management of Rivers Affected by Channel Regulation Projects. *Quat. Int.* **2015**, *384*, 180–185. [CrossRef]
74. Piggott, J.J.; Townsend, C.R.; Matthaei, C.D. Climate Warming and Agricultural Stressors Interact to Determine Stream Macroinvertebrate Community Dynamics. *Glob. Change Biol.* **2015**, *21*, 1887–1906. [CrossRef]
75. Verberk, W.C.E.P.; Hoefnagel, K.N.; Peralta-Maraver, I.; Floury, M.; Rezende, E.L. Long-term Forecast of Thermal Mortality with Climate Warming in Riverine Amphipods. *Glob. Change Biol.* **2023**, *29*, 5033–5043. [CrossRef]
76. Shi, X.; Liu, J.; You, X.; Bao, K.; Meng, B.; Chen, B. Evaluation of River Habitat Integrity Based on Benthic Macroinvertebrate-Based Multi-Metric Model. *Ecol. Model.* **2017**, *353*, 63–76. [CrossRef]
77. Sandin, L. Benthic Macroinvertebrates in Swedish Streams: Community Structure, Taxon Richness, and Environmental Relations. *Ecography* **2003**, *26*, 269–282. [CrossRef]
78. Jiang, X.; Xiong, J.; Qiu, J.; Wu, J.; Wang, J.; Xie, Z. Structure of Macroinvertebrate Communities in Relation to Environmental Variables in a Subtropical Asian River System. *Int. Rev. Hydrobiol.* **2010**, *95*, 42–57. [CrossRef]
79. Rico, A.; Van den Brink, P.J.; Leitner, P.; Graf, W.; Focks, A. Relative Influence of Chemical and Non-Chemical Stressors on Invertebrate Communities: A Case Study in the Danube River. *Sci. Total Environ.* **2016**, *571*, 1370–1382. [CrossRef] [PubMed]
80. Dalu, T.; Mwedzi, T.; Wasserman, R.J.; Madzivanzira, T.C.; Nhiwatiwa, T.; Cuthbert, R.N. Land Use Effects on Water Quality, Habitat, and Macroinvertebrate and Diatom Communities in African Highland Streams. *Sci. Total Environ.* **2022**, *846*, 157346. [CrossRef]
81. Shi, X.; Liu, J.; You, X.; Bao, K.; Meng, B. Shared Effects of Hydromorphological and Physico-Chemical Factors on Benthic Macroinvertebrate Integrity for Substrate Types. *Ecol. Indic.* **2019**, *105*, 406–414. [CrossRef]
82. He, K.; Chen, C.; He, Y.; Feng, X.; Tang, H.; Yang, L.; He, Y.; Lyu, W. Differential Responses of Macroinvertebrates and Phytoplankton to River Health Status in Different Habitats. *Aquat. Sci.* **2024**, *86*, 12. [CrossRef]
83. Kalogianni, E.; Vourka, A.; Karaouzas, I.; Vardakas, L.; Laschou, S.; Skoulikidis, N.T. Combined Effects of Water Stress and Pollution on Macroinvertebrate and Fish Assemblages in a Mediterranean Intermittent River. *Sci. Total Environ.* **2017**, *603*, 639–650. [CrossRef] [PubMed]