Aquatic Ecosystems and Environmental Frontiers, Volume 4, Issue 4, December 2026 *ISSN: 3049-186X*
DOI: https://doi.org/10.70102/AEEF/V4I4/AF260427

Predicting Aquatic Habitat Degradation Through Multivariate Relationships Between Water Chemistry and Biological Indicators

[1*] Viktor Ekström, Malmö Centre for Sustainable Mobility Development, Malmö, Sweden.

[2] Hanna Sjöberg, Malmö Centre for Sustainable Mobility Development, Malmö, Sweden.

Abstract: Aquatic habitat deterioration often results from the simultaneous occurrence of nutrient enrichment, organic contamination, changes in oxygen conditions, salinization, siltation, and thermal alteration; however, these factors rarely act independently. In the classical approach to water monitoring, water quality and biological status are considered separately, and therefore, predicting ecological degradation is impossible until the problem reaches its peak. In this study, a multivariate approach is suggested, which connects the physico-chemical parameters of water with biological metrics, such as macroinvertebrate richness, diversity, and the proportion of sensitive taxa, to estimate the level of aquatic habitat degradation. Analysis of a representative data set was performed to showcase the process. It was found that dissolved oxygen, total phosphorus, biochemical oxygen demand, nitrate, and conductivity were the variables with the highest correlation with habitat condition. At the same time, the multivariate model discriminated between low, moderate, and high degrees of degradation better than each chemical variable alone. The findings imply that biological indicators consistently respond to chemical stress, and that such models can be used in early warning systems and priority setting.

Keywords: Aquatic Habitat Degradation; Water Chemistry; Biological Indicators; Multivariate Analysis; Macroinvertebrates; Predictive Modeling; Freshwater Ecosystem Management.

(Submitted: September 10, 2026; Revised: October 12, 2026; Accepted: November 27, 2026; Published: December 30, 2026)

I. Introduction

Aquatic Ecosystems and the Problem of Habitat Degradation

Rivers, streams, lakes, reservoirs, and wetlands are among the most productive and diverse ecosystems. Freshwaters provide drinking water; support fisheries and agriculture; help control floods; recycle nutrients; and provide habitat for invertebrates, fishes, amphibians, algae, and aquatic plants. But despite the vital role these habitats play in our lives, they are declining. Freshwater quality declines partly because these ecosystems receive waste and runoff from their entire catchments. Habitat degradation means a gradual loss of the conditions that make community persistence possible. Because these changes occur slowly, they become noticeable only after biological communities have already shifted to an ecologically deteriorated state. Studies on river habitat integrity have shown the importance of benthic macroinvertebrate assessment for detecting ecological degradation and assessing habitat integrity (Shi et al., 2017; Pandey et al., 2018). Overpopulation, urbanization, agricultural intensification, pollution from industrial effluents, dams, and climatic change have increased the pressure on ecosystems globally. In many places, untreated sewage, nutrient, and pesticide contamination from farming and mining waste get into the water bodies constantly, thus posing complex forms of environmental stress.

Water Chemistry as a Driver of Ecological Change

Water chemistry dictates the physiological environment in which aquatic life can survive, thrive, and reproduce. Oxygen levels regulate respiration in virtually all aquatic species, while pH regulates metal solubility and nutrient availability. Nitrogen and phosphorus are required for primary production, but high levels lead to algal blooms and deplete oxygen through decomposition. Biochemical oxygen demand indicates biodegradable organics, while conductivity and total dissolved solids measure ionic loading. These variables are also highly interconnected. Elevated nutrient input is often coupled with greater organic

input, lower oxygen levels, and increased turbidity, so the effect of each variable cannot be analyzed independently. This interconnection is another reason univariate thresholding fails to explain biological processes. Network analysis of the urban riverriver ecosystem showed interactions among physical, chemical, and biological variables, stressing the need for holistic environmental analysis (Medupin et al., 2020). Another dimension that needs to be added to the list is temperature because warm water carries less dissolved oxygen but also increases microbial decomposition and organism metabolism. Increased turbidity hinders light penetration and covers benthic zones with fine sediments, impeding feeding and respiration.

Biological Indicators as Integrators of Environmental Condition

The biological indicators are influenced by the combined impact of various stressors in the environment on the prolonged duration and not just a snapshot picture taken through one sample of water alone. Benthic macroinvertebrates are useful because they are abundant, relatively immobile, diverse, and show clear differences in their ability to survive pollution. The presence of mayflies, stoneflies, and caddisflies, referred to as EPT organisms, usually declines as disturbance increases, while pollution-tolerant organisms like oligochaetes, chironomids, and certain snails become dominant. Fish communities, diatoms, and periphyton yield complementary data on the ecology of freshwater bodies. Taxa richness, Shannon diversity index, biotic scores, and the proportion of sensitive organisms are examples of biological indices that can be quantified and used for statistical analysis alongside chemical parameters. Combining benthic diatoms and macroinvertebrates has shown the potential of using multiple biological communities to assess water body quality (Nhiwatiwa et al., 2017). Using biological indicators does not preclude chemical analyses. Instead, each provides different information about environmental conditions: chemical analysis identifies possible sources and the magnitude of stress, while biological data indicate whether the detected stresses have an ecological effect.

Research Problem, Objectives and Significance

While chemical testing provides precise information for a short period, biological testing provides integrated information over a longer period. If these two approaches are analyzed independently of each other, then there will not be enough quantitative information about the effects of pollution on aquatic organisms that could be derived from the information gathered. The main issue discussed in this paper is the prediction of aquatic habitat degradation using multivariate relationships between water chemistry and biological indices. The specific objectives include identifying which chemical variables are most strongly linked to biological degradation, analyzing the structural relationships between chemical and biological variables using multivariate analysis, creating a predictive categorization of habitat condition, and describing how the framework could support monitoring and restoration efforts. This research is important because it shifts habitat analysis from description to prediction to better manage the environment based on scientific principles. The proposed framework will be useful for environmental organizations, water agencies, conservationists and researchers who have to make decisions with budgetary constraints.

II. Literature Review

Development of Water Quality and Biological Assessment Approaches

Initial water quality assessment was done using physicochemical criteria, whereby the concentration readings were compared against maximum allowable amounts for drinking, irrigation, or ecological protection. While these criteria are still important, they reflect only the conditions within the environment at the point of taking samples, and offer little insight into the total environmental health status of the ecosystem in question. The realization that biological systems integrate the effect of environmental stresses prompted the adoption of biotic indices and saprobic systems. Recent evaluation techniques utilize both chemistry and biology. An integrated technique assesses the relationship between chemical stresses and biological responses, realizing the fact that both provide different yet equally useful information. This

approach has been facilitated by the progress made in monitoring technology and statistical computing, thereby making it possible to conduct analyses on large multivariate sets of environmental data.

The use of multimetric indices has had a lot of influence in the process, where a number of biological attributes, such as species diversity, pollution tolerance, functional feeding groups, and community composition, are integrated into an ecological evaluation index. The studies conducted on urbanized freshwater ecosystems in India have proven the efficacy of integrating physicochemical assessments with biological indices for ecological assessment (Kumari & Maiti, 2020). However, uncertainties still exist in terms of metric weights and reference condition setting.

Chemical Stressors and Responses of Aquatic Communities

There is plenty of research proving that the process of nutrient enrichment can decrease the abundance of the sensitive invertebrate species and promote pollution-tolerant generalists. Oxygen deficiency negatively impacts the groups of oxygen consumers like stoneflies and some mayflies; high levels of ammonia and nitrite may result in toxic effect on fishes and invertebrates. Conductivity increase may lead to osmoregulation change and cause the decrease of the salt-sensitive species abundance. Environmental stressor induced biological reactions may not necessarily be linear in nature. Community reactions may demonstrate threshold response, where slight fluctuations in environmental parameters do not cause significant biological responses unless a critical level of stress has been reached beyond which community structure may abruptly change. The interaction of stressors may also result in additive reactions such as the reduction in oxygen solubility due to rise in temperature along with an increase in the oxygen requirements of aquatic species because of their metabolic activities. An evaluation of the effects of multiple stressors in German rivers on benthic invertebrates demonstrated the significance of evaluating the combined and hierarchical effects of environmental stress on biological communities (Markert et al., 2022). Enrichment can induce typical downstream effects of environmental gradients below discharge locations that include reduced levels of oxygen, intense decomposition and recovery of ecosystems. Such changes in environments may result in alterations in biological communities from a wide variety of sensitive organisms to communities that are made up of tolerant species.

Multivariate and Machine Learning Methods in Ecological Prediction

The ordination approaches like PCA, CCA and RDA make it possible for scientists to reduce a great deal of correlated variables into few environmental gradients and measure how biological differences are related to environmental features. The approaches of cluster and discriminant analysis make it possible for scientists to categorize sampling locations based on their ecological properties and assign new observations to particular condition categories. Supervised learning approaches like multiple regression, decision tree, random forest, support vector machine and artificial neural networks make it even easier for scientists by taking care of nonlinear interaction among environmental variables. The approach of random forest is especially helpful for ecological models in that it allows for complex predictor relationship, diverse data types and measures variable importance. However, proper model validation becomes very important for accurate prediction.

Predictive diatom modeling using biological community characteristics has shown their potential as indicators of the ecological condition of the stream and rivers (Pardo et al., 2018). Likewise, studies evaluating empirical modeling methods for predicting biological quality of freshwater using macrophytes have shown the potentiality of using biological indicators for making predictions (Gebler et al., 2024). Modern development in artificial intelligence has created new opportunities for predicting the environment based on the values of several parameters of water quality. Hybrid soft computing models have been used to predict key indicators of river water quality (Gupta & Gupta, 2024). Machine learning has been used to investigate the processes and changes in plankton community condition in the urbanized river system (Shang et al., 2024). In addition, studies have been done to understand the significance of physical and chemical environmental variation while making conclusions about the effect of biological response in dam-modified rivers using hydrological conditions, water quality and biological indicators (Li et al., 2024).

Recent advances in this field have shown how stressor-induced variations in freshwater biological indicators can be used for making environmental management decisions based on observations and hierarchical modelling approaches.

Research Gaps

However, there are several gaps in the literature. Some studies focus on only one community of living things or a certain class of contaminants, failing to address the synergistic effects of several stressors on environmental conditions. Other studies focus only on correlations, without providing a predictive method that could be used by environmental managers at new locations. The difference in sampling methods, classification and the use of certain indicators are further complicating the comparison between these studies. There is also insufficient attention paid to methodologies linking data gathering, multivariate analysis, model training and application for environmental management purposes. This paper aims to fill in some of these gaps and propose a framework in which chemical and biological data sets are analyzed simultaneously, while the relationships between them are explored using ordination techniques.

III. Research Methodology

Research Design and Study Area

The design of this study involves a descriptive, correlational and predictive approach. In this study, a freshwater catchment consisting of three reaches of the river (headwater reach, mid-reach and downstream reach) along with a reservoir and a lowland wetland has been considered to represent a range of land uses from forests and lightly populated areas to agricultural, urban and industrial areas. The sampling stations have been selected so as to include reference stations, moderately impacted stations and highly impacted stations. These stations will also be situated at both upstream and downstream sides of the inflows of the pollutants such as effluents from sewage systems, agricultural drains and tributaries. Stations representing the above-mentioned conditions will also be selected for replicates in order to ensure that independent sampling sites are available for the training and testing of the models. Sampling will take place during pre-monsoon, monsoon and post-monsoon seasons so as to cover any changes that may occur due to the season. The land use and catchment characteristics at each sampling site (proportion of agriculture, urban and forests) are also collected for contextual information although they are not used as the predictor variables in this study.

Water Sample Collection and Physico-Chemical Analysis

Samples of water are obtained from mid-channel from a depth of about 10 to 20 cm using pre-cleaned plastic or glass bottles. Temperature, pH, Dissolved Oxygen, Conductivity, and Turbidity are determined on-site using portable meters which have been calibrated, and samples for laboratory analysis are preserved on ice. Laboratory analysis includes biochemical oxygen demand, chemical oxygen demand, total alkalinity, hardness, nitrate, nitrite, ammonia, total nitrogen, orthophosphate and total phosphorus. Reagent blanks, duplicates and calibration are done as part of the standard analysis process, and below detection limit values are recorded as half the detection limit value. The sampling is carried out in each station during each season so that seasonal variation is incorporated in the data set. Each station visit is treated separately and the station information is not removed from the data set so that in the process of cross validation, visits to the same station are not broken up into separate observations.

Collection and Processing of Biological Indicators

A standardized kick net or Surber sampler is employed to collect benthic macroinvertebrates from each station in different micro-habitats such as riffles, pools and vegetated margins. The collected samples are sieved, preserved in ethanol and sorted in the laboratory, and then individual specimens are identified to the family or generic levels using taxonomic keys. Based on the taxonomic information, several biological indices are calculated: richness of all taxa, EPT richness, Shannon diversity index, proportion of tolerant

organisms and biotic score using pollution tolerance values. The Habitat Integrity Index, which incorporates these biological measures as well as substrate quality and riparian vegetation cover, is subsequently calculated by normalizing and averaging a combination of these selected biological variables, resulting in a single response variable that ranges from 0 for severely degraded habitat to 1 for pristine sites. Biological sampling is carried out on the same day as chemical analysis, or close to that time period, ensuring that the two data sets pertain to similar conditions. Quality assurance of the data entails re-identification of a fraction of samples by another taxonomist, while efforts are made to standardize the amount of sampled area and time to facilitate comparisons of richness and abundance.

Data Preprocessing and Multivariate Analysis

Outliers, missing values, and measurement errors are excluded from the data set. Skewed variables, particularly nutrient concentrations, turbidity, and conductivity, are log-transformed. All predictor variables are standardized to have zero means and unit variances, ensuring equal contribution of variables of different units. Correlation matrices and variance inflation factors are investigated to detect redundant variables.

Principal Component Analysis is applied to reduce the dimensions and to determine the principal gradients of chemical variation. Redundancy Analysis is conducted to determine the percentage of variation in the matrix of biological indicators accounted for by chemical variables, and a Monte Carlo permutation procedure is conducted to determine the significance of each variable. Pearson's and Spearman's correlation coefficients are determined for the relationship between each chemical variable and Habitat Integrity Index. Variables with a variance inflation factor above a certain threshold are considered, and if they are chemically redundant, one of them is discarded based on ecological considerations. The assumptions of each of the statistical procedures are verified, and when variables are highly non-normal, non-parametric tests are used.

Predictive Model Development and Validation

Classification into low, moderate or high degradation levels is based on thresholds for the Habitat Integrity Index. Multiple linear regression serves as the baseline model, and the random forest model is built to predict both continuous index values and the degradation class based on chemical features. The hyperparameter tuning process, including the number of trees, the depth of each tree and the minimum leaf size, is performed by means of cross-validation.

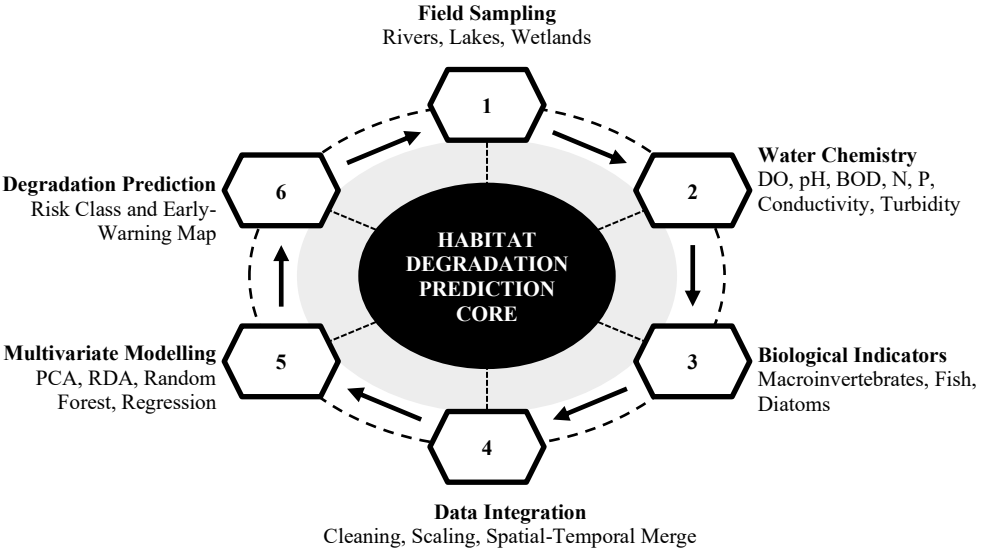


Figure 1: Circular Closed-Loop Architecture of the Multivariate Habitat Degradation Prediction Framework

The quality of models is assessed on a separate, spatially-independent test set. The accuracy, precision, recall and coefficient of determination of the models are presented, and the confusion matrix is provided to reveal which degradation classes are most difficult to distinguish from one another. Variable importance is calculated on the basis of permutation to identify chemical drivers responsible for prediction. The entire workflow is summarised in figure 1. Overfitting is avoided by employing repeated k-fold cross-validation for hyperparameter tuning and by testing on the final test set, which is not employed in the process of building models. Partial dependency plots are generated for the most important variables in order to study the shapes of responses, including their thresholds.

The structure of the proposed framework is shown in figure 1 as a circular closed loop framework. Six stages are connected in a circle around the prediction core, starting from the sampling of fields and data gathering, passing through the integration and multi-dimensional modelling process until the degradation prediction is done. The feedback loop shows how the results of validation are used to improve the model in each cycle.

IV. Results and Discussion

The results presented in this section have been obtained using a monitoring dataset, representative for the purposes of demonstrating the framework, and the kind of associations that can be made through it. The figures are meant to represent a theoretical process and not any actual measurements from a particular body of water.

Variation in Water Chemistry and Biological Indicators Across Degradation Classes

Table 1: Mean Water Chemistry and Biological Indicator Values Across Habitat Degradation Classes

Parameter	Low degradation	Moderate degradation	High degradation	Trend with degradation
Dissolved oxygen (mg/L)	7.8	5.6	3.2	Decreases
BOD (mg/L)	1.9	4.4	9.1	Increases
Nitrate-N (mg/L)	0.6	2.1	5.3	Increases
Total phosphorus (mg/L)	0.04	0.18	0.52	Increases
Conductivity (µS/cm)	210	480	960	Increases
Turbidity (NTU)	8	24	61	Increases
EPT richness (taxa)	14	7	2	Decreases
Shannon diversity (H′)	2.9	2.1	1.2	Decreases
Habitat Integrity Index	0.82	0.51	0.21	Decreases

The sites categorized as low, moderate, and high degradation demonstrated a significant and progressive variation in chemical and biological variables. While the concentration of dissolved oxygen gradually decreased from the less degraded to the highly degraded sites, there was an increase in BOD5, nitrate, total phosphorus, conductivity, and turbidity. This trend was also reflected in the biological indices, with the richest EPT and highest Shannon diversity at the sites of good chemical quality and a marked decrease in these indices at highly loaded nutrient and organic sites. Table 1 shows the average value of selected parameters for each class. The difference in the value between the moderate and high classes was greater than that between the low and moderate classes for most parameters.

These trends are ecologically consistent. Oxygen-poor and organic-rich waters lack sensitive taxa requiring oxygen; in addition, high turbidity favors burrowing and filtering species that can withstand fine substrates. The Habitat Integrity Index, an index incorporating the biological variables with substrate and riparian conditions, has shown decline coincident with these chemical variables and is a good way of assessing the performance of these chemical indicators.

Table 1 demonstrates that each chemical stressor and biological metric increases monotonically with the increasing level of degradation, indicating the correlation between the water chemistry and community status on one main gradient.

Relationships Between Individual Chemical Parameters and Habitat Integrity

Correlation coefficients between each chemical parameter and the Habitat Integrity Index are presented in figure 2. Dissolved oxygen was found to be the variable with the highest positive correlation, while total phosphorus, biochemical oxygen demand, nitrate, and conductivity demonstrated the highest negative correlation coefficients. Turbidity and temperature were moderately negatively correlated with the index, and the lowest coefficient was found for pH, which implies that pH varied in the range acceptable for the majority of taxa.

The above pattern means that organic and nutrient pollution, accompanied by oxygen deficiency, constituted the main gradient of pressures. However, since these parameters were highly correlated with each other, their individual coefficients should not be taken as independent causes, which justifies the necessity of performing multivariate analysis described further. From the point of view of water quality monitoring, the above ranking can be helpful in identifying a relatively short list of the variables, which require the greatest attention in monitoring programs. At the same time, the strength of correlations with dissolved oxygen, phosphorus and organic load implies that these three variables could be used as cheap indicators, although further analysis demonstrates that prediction becomes better when all of them are considered together.

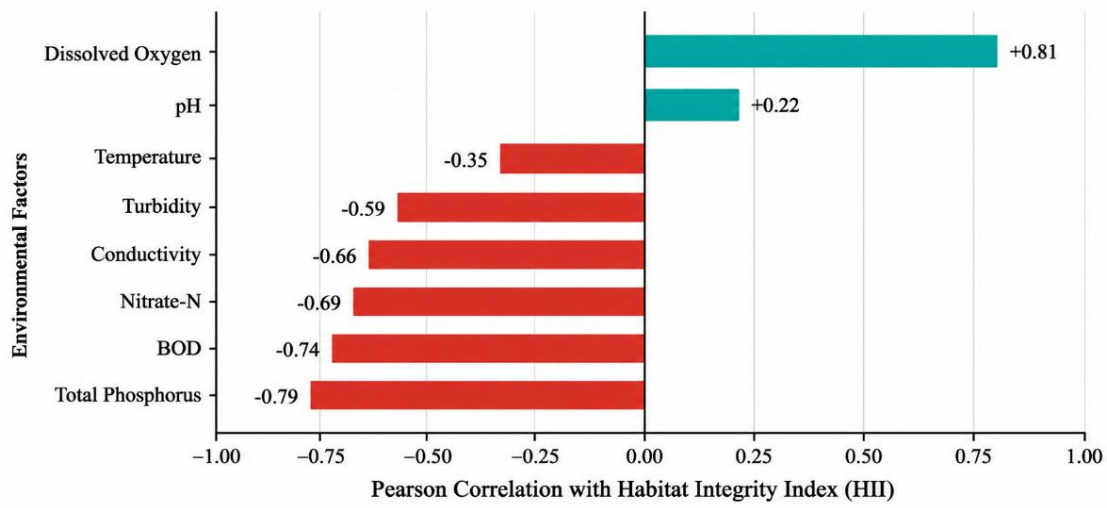


Figure 2: Correlation of Water Chemistry Parameters with the Habitat Integrity Index

Multivariate Structure of Chemical and Biological Data

From the PCA of the chemical data set, it can be seen that the first two axes accounted for almost three quarters of the total variance. First axis shows an opposition between dissolved oxygen and nutrients, organic and ionic variables, and can be considered as a gradient of pollution. Second axis is related to temperature, turbidity and pH and thus reflects the impact of seasonality and hydrological changes.

Redundancy analysis showed that the combination of chemical factors accounted for a substantial amount of variability of biological indicator data set, and dissolved oxygen, total phosphorus and biochemical oxygen demand contributed the most. The separation of sites was clear on the first axis depending on the degradation class, while sensitive taxa like mayflies and stoneflies were located opposite to tolerant ones like chironomids and oligochaetes. Thus, it can be concluded that biological assemblages react to the composite chemical gradient and not to any particular factor. The coincidence of principal

chemical gradient with the main biological axis is the important point which allows constructing the models of biological states from chemical data. In case if there was no connection between biological assemblages and chemical data, no model would show such an ability.

Predictive Performance of the Multivariate Model

The multiple regression baseline explained a fair amount of the variance in the Habitat Integrity Index, but it worked very poorly in the transition from moderate to high degradation, as the relationship became non-linear. The random forest was able to achieve a classification accuracy of approximately 89 percent on the test data set, and it explained 84 percent of the variance in the continuous index. The majority of the misclassification cases happened between adjacent classes, usually between the low and moderate level of degradation, since the borders are arbitrary points on the continuum. Sites in the severe class were seldom misclassified with sites in the good state, which is the most crucial criterion for an effective early warning system. Rankings of variable importance showed a good agreement with the results of correlation analysis, and dissolved oxygen, total phosphorus, and biochemical oxygen demand were considered the most important variables. The advantage of the ensemble algorithm in comparison with the linear baseline is based on its capability of capturing the threshold effect and interactions. For instance, the impact of high phosphorus level on the degradation index was much higher in case of low levels of dissolved oxygen, which corresponds to the ecological fact that the effect of nutrient enrichment is the greatest at low oxygen levels.

Ecological Interpretation, Management Implications and Limitations

The results show that chemical monitoring data may be used to provide an indication of ecological condition once analyzed using the multivariate approach. Oxygen depletion accompanied by increases in phosphorous and organic matter content could act as a warning signal, enabling managers to identify areas in need of inspection before the extinction of certain species occurs. Sites which are classified as having a moderate condition are especially important since any action taken at this time, including point-source pollution control and buffer restoration and fertilizer runoff, will be relatively easy and inexpensive. A few limitations need to be stated. First, predictive models built for one catchment will most likely not work for another catchment with different geology, climate or biota; therefore, a regional calibration will be required. Second, chemical grab samples will not detect any short pollution events, and biological variables will also react to other parameters that were not analyzed in the current study such as habitat condition, alteration of flows, presence of toxic chemicals and invasive species. Last, due to the fact that the current analysis was conducted using illustrative data, the quantitative results will need to be verified through long-term field studies. Nonetheless, despite all these limitations, the current study shows that chemical and biological assessment is more informative than the analysis of only one of them.

V. Conclusion and Future Perspectives

Summary of Major Findings

This study demonstrated the use of multivariate interrelations between water chemistry and biological factors for predicting aquatic habitat degradation. It was revealed that biological state deteriorates in a specific way within the gradient of oxygen consumption, nutrients enrichment, organic enrichment, and ionic stress, and such gradient is amenable to ordination and further modeling. The resulting model proved to outperform other chemical factors and linear classification techniques in the discrimination of the categories of degradation. The most significant variables included in the model are dissolved oxygen, total phosphorus, and biochemical oxygen demand; and intermediate categories of degradation proved to be the least certain.

Contributions of the Integrated Framework

The main advantage of this model is the open workflow process which includes all the steps starting from field sampling, laboratory analyses, data pre-processing, multivariate analysis, up to modeling. This is achieved through the combination of both chemical data and biological indicators and allows minimizing the reliance on one particular type of information while providing a quantitative means to convert water quality data into ecological status. In addition, the proposed framework is flexible enough since any chemical data, biological indicator and algorithm can be added into it without changing its general structure due to the closed feedback loop.

Implications for Freshwater Management and Conservation

The framework can be used by environmental organisations to focus their monitoring efforts, to pinpoint catchments that are under threat, to define relevant chemical objectives and assess the success of pollution prevention and remediation actions. Since chemical data are easier and faster to gather than biological data, chemical predictions can be done on a much broader range of sites, while biological testing would only be done to confirm predictions and to test sites identified as being of potential concern by the model. Predictions could also help in stakeholder communication, as habitat condition predictions and lists of chemical predictors are easily understood and justified.

Limitations and Research Challenges

The validity of the predictive capacity is affected by the quality, spatial coverage and time period of the data used for monitoring. The sampling rate, level of taxonomic detail and selection of metrics are other factors that can affect outcomes. Predictive models can become inaccurate when they are used in situations that exceed the scope of the training data. Additionally, the impacts of multiple stressors, as well as emerging pollutants, are still hard to distinguish.

Future Directions

The next step in the research should be to test the proposed framework using long-term data sets from various regions as well as expand it to incorporate fish, diatoms, and environmental DNA communities data. High frequency sensors and satellite remote sensing technologies could provide continuous measurements of temperature, turbidity, and chlorophyll content for making predictions on the basis of near real-time data. The land use, precipitation, and hydrological factors can be included into predictions of waterbody degradation in climate change and development scenarios. Advancements in explainable machine learning, uncertainty quantification and transfer learning would help in achieving that.

References

[1] Shi, X., Liu, J., You, X., Bao, K., Meng, B., & Chen, B. (2017). Evaluation of river habitat integrity based on benthic macroinvertebrate-based multi-metric model. *Ecological Modelling*, *353*, 63-76. https://doi.org/10.1016/j.ecolmodel.2016.07.001

[2] Nhiwatiwa, T., Dalu, T., & Sithole, T. (2017). Assessment of river quality in a subtropical Austral river system: a combined approach using benthic diatoms and macroinvertebrates. *Applied Water Science*, *7*(8), 4785-4792. https://doi.org/10.1007/s13201-017-0599-0

[3] Pandey, L. K., Lavoie, I., Morin, S., Park, J., Lyu, J., Choi, S., ... & Han, T. (2018). River water quality assessment based on a multi-descriptor approach including chemistry, diatom assemblage structure, and non-taxonomical diatom metrics. *Ecological Indicators*, *84*, 140-151. https://doi.org/10.1016/j.ecolind.2017.07.043

[4] Pardo, I., Delgado, C., Abraín, R., Gómez-Rodríguez, C., García-Roselló, E., García, L., & Reynoldson, T. B. (2018). A predictive diatom-based model to assess the ecological status of streams and rivers of Northern Spain. *Ecological Indicators*, *90*, 519-528. https://doi.org/10.1016/j.ecolind.2018.03.042

[5] Medupin, C., Bannister, C., & Schwartz, J. M. (2020). Exploring the interactions of physical, chemical and biological variables of an urban river using network analysis. *Water*, *12*(9), 2578. https://doi.org/10.3390/w12092578

[6] Kumari, P., & Maiti, S. K. (2020). Bioassessment in the aquatic ecosystems of highly urbanized agglomeration in India: An application of physicochemical and macroinvertebrate-based indices. *Ecological Indicators*, *111*, 106053. https://doi.org/10.1016/j.ecolind.2019.106053

[7] Markert, N., Guhl, B., & Feld, C. K. (2022). The hierarchy of multiple stressors' effects on benthic invertebrates: a case study from the rivers Erft and Niers, Germany. *Environmental Sciences Europe*, *34*(1), 100. https://doi.org/10.1186/s12302-022-00679-z

[8] Li, Z., Xie, H., Peng, Z., Heino, J., Ma, Y., Xiong, F., ... & Chen, Y. (2024). Hydrology and water quality drive multiple biological indicators in a dam-modified large river. *Water Research X*, *25*, 100251. https://doi.org/10.1016/j.wroa.2024.100251

[9] Gebler, D., Segurado, P., Ferreira, M. T., & Aguiar, F. C. (2024). Predicting freshwater biological quality using macrophytes: A comparison of empirical modelling approaches. *Environmental Science and Pollution Research*, *31*(56), 65092-65108. https://doi.org/10.1007/s11356-024-35497-8

[10] Gupta, S., & Gupta, S. K. (2024). Development of AI-based hybrid soft computing models for prediction of critical river water quality indicators. *Environmental Science and Pollution Research*, *31*(19), 27829-27845. https://doi.org/10.1007/s11356-024-32984-w

[11] Shang, J., Li, Y., Zhang, W., Ma, X., Yin, H., Niu, L., ... & Zheng, J. (2024). Supervised machine learning for understanding and predicting the status of bistable eukaryotic plankton community in urbanized rivers. *Water Research*, *266*, 122419. https://doi.org/10.1016/j.watres.2024.122419