Ecological Informatics 90 (2025) 103251

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Combining environmental DNA and remote sensing variables to model fish biodiversity in tropical river ecosystems[☆]

Robin Bauknecht [a,*], Loïc Pellissier [a,b], Sébastien Brosse [c], Vincent Prié [d,f,g], Manuel Lopes-Lima [d,f], Pedro Beja [d,e,f], Monika K. Goralczyk [a,b], Andrea Polanco Fernandez [h], Jorge A. Moreno-Tilano [i,j], Rafik Neme [i,j], Mailyn A. Gonzalez [k], Shuo Zong [a,b,**]

[a] *Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland*
[b] *Land Change Science Research Unit, Swiss Federal Institute for Forest, Snow and Landscape Research, WSL Birmensdorf, Switzerland*
[c] *Centre de Recherches sur la Biodiversité et l'Environnement, CRBE UMR5300, Université Paul Sabatier, IRD, CNRS, INP 118 route de Narbonne, 31062 Toulouse, France*
[d] *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal*
[e] *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal*
[f] *BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal*
[g] *SPYGEN, 17, rue du Lac Saint-André - CS 20274, 73375 Le Bourget-du-Lac, France*
[h] *Fundacion Biodiversa Colombia, Bogota, Colombia*
[i] *Department of Chemistry and Biology, Universidad del Norte, Barranquilla, Colombia*
[j] *Max Planck Partner Group for Genomics and Biodiversity of the Colombian Caribbean, Colombia*
[k] *Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia*

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.15869405

Keywords:
Environmental DNA
Remote sensing
Ecological modeling
Biodiversity monitoring
Aquatic species detection
Tropical river ecosystems

ABSTRACT

Tropical river ecosystems face substantial threats, leading to a sharp decline in their biodiversity. High-resolution data on the spatial distribution of biodiversity is essential for devising effective conservation strategies. However, biodiversity information is limited because traditional assessment methods often face challenges in these vast, inaccessible environments. Here, we aim to assess whether combining large-scale environmental DNA (eDNA) data with environmental variables generated from remote sensing images in machine learning models can overcome this limitation. We used a fish eDNA dataset of 264 samples collected from major tropical rivers—the Casamance, Cuando, Cunene, Okavango, and Zambezi (Africa); the Magdalena, Maroni, and Oyapock (South America); and the Kinabatangan (Southeast Asia)—together with aquatic and terrestrial variables derived from remote sensing imagery. Based on this data, we constructed both river-specific and multi-river Random Forest models to predict fish species richness and the Shannon diversity index. The models exhibited a good fit to the data, indicating the suitability of variables in capturing the determinants of fish biodiversity in these rivers. Moreover, the models effectively predicted the metrics during cross-validation, underscoring their utility in generating biodiversity maps along large tropical rivers. Although predictions for unencountered rivers remain challenging, the models are able to capture large-scale patterns. With further refinement and expansion through additional data, this integrated approach holds promise for generating biodiversity insights without extensive on-site sampling requirements. Our study highlights the potential of combining eDNA with remote sensing variables to model biodiversity patterns in tropical river ecosystems.

[☆] This work was supported by a China Scholarship Council grant awarded to S. Z. This project has also received support from the project NORTE-01-0145-FEDER-000046 under the Norte Portugal Regional Operational Programme (NORTE2020), through the European Regional Development Fund (ERDF) and the Portugal 2020 Partnership Agreement. Additionally, M. L.-L. was funded by FCT - Fundação para a Ciência e Tecnologia [contract 2020.03608.CEECIND]. The Maroni River sampling and eDNA analyses were funded by the Office de l'Eau de la Guyane. SPYGEN provided financial support for all eDNA analyses presented in this study, within the framework of the Vigilife project.

[*] Corresponding author.

[**] Corresponding author at: Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland.

E-mail addresses: rbauknecht@ethz.ch (R. Bauknecht), shuo.zong@usys.ethz.ch (S. Zong).

https://doi.org/10.1016/j.ecoinf.2025.103251
Received 5 August 2024; Received in revised form 29 May 2025; Accepted 30 May 2025

1. Introduction

Tropical rivers are some of the most diverse ecosystems on Earth, characterized by a remarkable array of species coexisting in these vibrant aquatic environments (Dudgeon, 2000; Encalada et al., 2019). In particular, they provide habitat for a large number of fish species, with 51% of all freshwater fish species occurring in six major tropical river basins (Val et al., 2022). By sustaining fisheries, allowing transport, and serving as a source of drinking water, rivers further play a vital role in supporting the livelihoods of local communities residing along their shores (Lázaro-Vázquez et al., 2018; McIntyre et al., 2016). However, the biodiversity in rivers is encountering considerable threats as various as the construction of dams, land use alterations, and the introduction of invasive species (Darwall et al., 2018; Carrizo et al., 2017; Salgado et al., 2022). Comprehensive understanding is essential for devising impactful, data-driven conservation strategies and monitoring programs to effectively reverse the biodiversity decline (Darwall et al., 2018; Miqueleiz et al., 2023).

Environmental DNA (eDNA) can enhance the capacity to quickly survey and assess biodiversity in freshwater ecosystems (Baird and Hajibabaei, 2012; Pawlowski et al., 2021). This complex mixture of intracellular DNA from living cells and extracellular DNA from sources such as skin, hair, urine, feces, or carcasses is directly extracted from environmental samples (Pawlowski et al., 2020). Metabarcoding, which combines high-throughput DNA sequencing with taxonomic reference libraries, then allows for the rapid and comprehensive identification of the taxa from which the DNA originates, thereby facilitating ecosystem assessment across large spatial scales (Deiner et al., 2016, 2017; Ruppert et al., 2019; Cilleros et al., 2019). Furthermore, eDNA stands out as a cost-effective method with a high detection capacity (Boivin-Delisle et al., 2021; Polanco Fernández et al., 2021). Alongside the methodological refinement, eDNA is progressively gaining prominence in assessing large-scale freshwater biodiversity (Altermatt et al., 2020; Lyet et al., 2021; Prié et al., 2023). Recent studies have demonstrated the effectiveness of eDNA in studying biodiversity within large river systems (Pont et al., 2018; Cheng et al., 2023; Cantera et al., 2022; Coutant et al., 2023). However, most investigations rely on data collected at discrete sampling locations, which are often separated by considerable spatial gaps—sometimes spanning tens of kilometers. This reliance on isolated sampling points limits the ability to capture continuous spatial patterns of biodiversity across entire river systems.

Satellite remote sensing (hereafter, remote sensing) provides regular, gap-free observations of the Earth's surface over time (Ozesmi and Bauer, 2002), which can serve as a valuable tool for acquiring information on the environmental conditions and habitat structures of both the aquatic and terrestrial surrounding of eDNA sampling areas (Zong et al., 2023). Therefore, integrating remote sensing data with eDNA sampling enables a more comprehensive understanding of riverine ecosystems (Domisch et al., 2015; Revenga et al., 2005; Campbell et al., 2011). Remote sensing provides high spatial and temporal resolution observation coverage, capturing a wide range of ecological properties, including vegetation indices (Sims et al., 2006; Crippen, 1990), surface temperature (Malakar et al., 2018; Vanhellemont, 2020a), water quality parameters (Olmanson et al., 2008), and hydrological characteristics (Schmugge et al., 2002). This capability facilitates detailed, continuous mapping and monitoring of habitat conditions across broad spatial and temporal scales, which is particularly valuable for assessing the heterogeneity of riverine landscapes (Hugue et al., 2016; Mertes et al., 1995). Thereby, remote sensing enables the modeling and interpolation of eDNA data, bridging the gap between discrete sampling and broader ecological patterns. This approach has the potential to provide critical ecological insights and improve biodiversity monitoring in complex riverine landscapes.

Fish are recognized as indicators for monitoring the health of large river ecosystems (Schiemer, 2000; Schmutz et al., 2007), making them an appropriate focal point for evaluating the functionality and reliability of the integrated approach. Fish habitat requirements are influenced by a spectrum of factors encompassing both aquatic and terrestrial realms (Dubey et al., 2012; Humpl and Pivnička, 2006). In terms of water variables, fish occurrence is expected to be influenced by water quality, water clarity, and water temperature (Rashid et al., 2018). The quality of water, characterized by factors such as nutrient levels and pollutant concentrations, impacts the availability of resources and the overall health of the aquatic habitat (Li et al., 2012; Sutela et al., 2010). Water clarity affects visibility and thus influences predator–prey interactions (Offem et al., 2009). Water temperature, a critical determinant of fish physiology, influences metabolic rates and reproductive patterns (Marshall and Elliott, 1998; Wolter, 2007). In terms of terrestrial variables, vegetation greenness, and canopy height collectively serve as critical indicators of riparian health, influencing the availability of terrestrial insects and fruits, providing essential shade for fish habitats, and creating diverse microhabitats that influence the distribution and behavior of fish species (Cetra and Petrere, 2007; Beltrão et al., 2009; Montag et al., 2019). Human modifications to the landscape, such as urbanization and agricultural activities, introduce alterations to the physical and chemical properties of the ecosystem (Leitão et al., 2018). In addition, human landscape use can also indicate potential anthropogenic pressures, including fishing practices (Ferguson et al., 2013) and pollution (Ngoye and Machiwa, 2004; Su et al., 2021). Finally, elevation, a topographical factor, influences the overall flow and structure of river systems, thereby affecting fish habitats (Carvajal-Quintero et al., 2015; Picado Barboza and Umaña Villalobos, 2018). All these environmental variables can be approximated using satellite images (Bergen et al., 2009; Drusch et al., 2012; Turner et al., 2003), and their utility has previously been demonstrated in fish species distribution modeling (Zong et al., 2023).

This study explores the potential of integrating eDNA metabarcoding with remote sensing data to model fish biodiversity in tropical river systems. Given the challenges of traditional biodiversity assessments in these vast and often inaccessible environments, this approach offers a promising means of generating information. We hypothesize that fish distributions within individual rivers are strongly influenced by environmental variables in both the aquatic and terrestrial realms, and consequently expect that models combining eDNA and remote sensing data should perform well at predicting biodiversity patterns within a single river. However, we also anticipate significant differences in baseline biodiversity among different rivers, driven by variations in regional species pools and biogeographic factors. This cross-river variability may present greater challenges for model performance when applied across river systems. However, understanding the extent of this predictive capacity is crucial for generating comprehensive biodiversity information in tropical freshwater ecosystems. To test these hypotheses, we analyze biodiversity metrics — species richness and Shannon index — derived from eDNA samples collected across 264 sites in nine major tropical rivers spanning Africa, South-East Asia, and South America. Using remote sensing data, we extract key environmental variables representing aquatic and terrestrial habitat conditions and then trained Random forest models to assess the relationships between these variables and biodiversity metrics. Specifically, the study seeks to address the following research questions:

1. To what extent is the variation in fish diversity in tropical rivers driven by local habitat features versus differences among river systems?
2. How effectively can models combining eDNA and remote sensing data capture the patterns of fish diversity in tropical river ecosystems?
3. How accurately can the integrated models predict the variation in diversity within and across different tropical rivers?

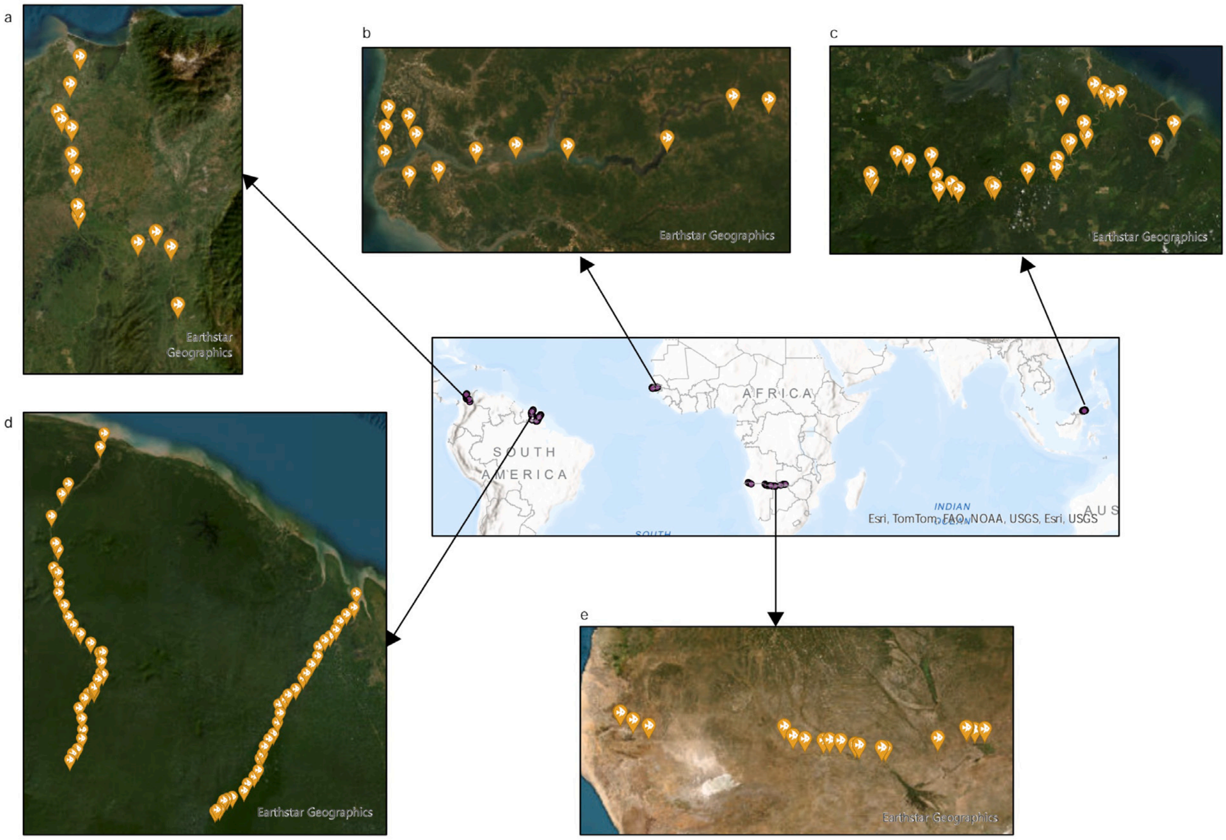

Fig. 1. Map showing the locations of 264 environmental DNA (eDNA) sampling sites distributed across nine major tropical rivers: the Magdalena (a), Casamance (b), Kinabatangan (c), Maroni and Oyapock (d), and the Cuando, Cunene, Okavango, and Zambezi (e).

2. Methods

2.1. Environmental DNA sample sites

This study utilized a comprehensive eDNA dataset from 264 samples collected in nine tropical rivers spanning three continents—the Casamance (Senegal), Cuando (Namibia), Cunene (Angola/Namibia), Kinabatangan (Malaysia), Magdalena (Colombia), Maroni (French Guiana), Oyapock (French Guiana), Okavango (Angola/Namibia), and Zambezi (Namibia/Zambia) river systems (Fig. 1). The selected rivers encompass varied tropical ecosystems, allowing for a comprehensive evaluation of the approach across multiple habitat types. The eDNA data for the Maroni and Oyapock rivers in French Guiana were sourced from an existing dataset (Cantera et al., 2022) and included 36 sites on the Maroni River sampled in 2017, each with two replicates (68 samples), and 38 sites on the Oyapock River sampled in 2018, each with two replicates (74 samples). For the remaining rivers, the data were collected during various sampling campaigns. Specifically, the dataset included 13 sites on the Casamance River sampled in 2018, each with two replicates (26 samples); 2 sites on the Cuando River sampled in 2022, each with two replicates (4 samples); 5 sites on the Cunene River sampled in 2022, each with two replicates (10 samples); 25 sites on the Kinabatangan River sampled in 2019, with 11 sites having two replicates (36 samples); 13 sites on the Magdalena River sampled in 2022, with 10 sites having two replicates (23 samples); 8 sites on the Okavango River sampled in 2022, with 7 sites having two replicates (15 samples); and 4 sites on the Zambezi River sampled in 2022, each with two replicates (8 samples). Due to the small sample size and close proximity of the Cuando, Cunene, Okavango, and Zambezi rivers in Southern Africa, we grouped them together in the analysis and will hereafter refer to them collectively as the Okavango. The same sampling methodologies were employed in the different rivers which allows for comparisons between the diverse datasets used in this analysis (A.1).

2.2. Environmental DNA processing

2.2.1. Laboratory processing

The water samples of all rivers were subjected to DNA extraction, amplification (12 PCR replicates per field sample), and high-throughput sequencing, following the established standard methodologies of Spygen (Cantera et al., 2022). Consistent across all rivers, for DNA amplification, the mitochondrial 12S rRNA gene fragment was targeted, utilizing "teleo" primers (forward primer-ACACCGCCCGTCACTCT, reverse primer-CTTCCGGTACACTTACCATG) known for their relevant performance in detecting bony fish species from eDNA (Valentini et al., 2016).

2.2.2. Bioinformatic analysis

A specialized bioinformatic pipeline was employed to analyze the high-throughput sequencing data derived from the samples (Marques et al., 2020). Initially, paired-end reads were merged using VSEARCH (Rognes et al., 2016), and a demultiplexing process was performed on the PCR replicate level. Subsequent steps included primer trimming using Cutadapt (Martin, 2011) and the dereplication of sequences using VSEARCH (Rognes et al., 2016) to facilitate redundancy removal. Molecular Operational Taxonomic Units (MOTUs) were defined using the SWARM clustering algorithm (Mahé et al., 2014). MOTUs are clusters of highly similar DNA sequences grouped based on genetic similarity, effectively serving as proxies for species. This method proves advantageous in tropical river ecosystems, where biodiversity is high, yet reference databases often lack comprehensive information (Marques et al., 2021; Polanco Fernández et al., 2021). Previous research demonstrated the effectiveness and reliability of MOTUs in estimating species diversity based on the "teleo" primer in the Rhone River (Marques et al., 2020). After clustering the MOTUs, chimeric sequences were removed, and MOTUs with fewer than ten reads were excluded to mitigate the inclusion of MOTUs originating from PCR errors. Taxonomic assignments were then achieved using the ecotag algorithm implemented in the Obitools toolkit (Boyer et al., 2016), providing a taxonomic context to the identified MOTUs. Therefore, as reference database, a combination of the European Nucleotide Archive (ENA) reference database (release 143, March 2020) and an internal custom reference database containing approximately 800 sequences (Mathon et al., 2023) was used. Following the taxonomic assignment, tag-jumps (Schnell et al., 2015) were corrected by removing sequences with an abundance frequency of less than 0.001 per taxon/MOTU and per library. When the reference database has limited species coverage, the ecotag algorithm (Boyer et al., 2016) may assign sequences to genus or family levels, even with low similarity. To address this, the assignments were refined by assigning species level for sequences matching at >98% similarity, genus level for sequences matching at 96%–98% similarity, family level for sequences matching at 90%–96% similarity, and order or higher level for sequences matching at less than 90% similarity. These adjustments were made to improve the accuracy of taxonomic assignments, aligning with the criteria proposed by Marques et al. (2020). Finally, the LULU algorithm (Frøslev et al., 2017) was implemented using an 84% identity threshold to clean MOTUs flagged as erroneous based on sequence identity between MOTUs, along with considerations of patterns of co-occurrence as well as abundances.

2.3. Biodiversity metrics

In the biodiversity computations, the unit employed is not a traditional species but the MOTUs generated by the SWARM algorithm. Throughout this manuscript, references to 'species' should therefore be understood as referring to MOTUs. This choice stems from the realization that the depth of available reference data in tropical rivers is limited, making it challenging to assign a substantial number of sequences to species levels (Marques et al., 2020). MOTUs serve as a pragmatic approach, allowing us to capture the diversity of fish populations based on genetic similarity rather than relying solely on species-level classifications in a system with limited reference data base depth (Marques et al., 2020). Biodiversity assessments were conducted by using two key metrics: species richness and the Shannon index. Species richness is a metric that quantifies the number of fish species present within a given sample or location (Lamb et al., 2009). A higher species richness indicates a more diverse fish community, reflecting a greater variety of ecological niches and roles within the ecosystem. The Shannon index (Margalef, 1958), on the other hand, considers both species richness and the evenness of species distribution (Lamb et al., 2009). A higher Shannon index reflects a community where species are more evenly distributed, indicating a balanced and resilient ecosystem (Spellerberg, 2008). The abundance in the Shannon index was estimated by the number of reads found for a particular MOTU cluster. Whilst the number of reads must not always precisely depict the number of individuals of a given species, it was previously shown that eDNA can be used to broadly quantify relative fish abundances (Lacoursière-Roussel et al., 2016; Spear et al., 2021; Yates et al., 2019). In particular, the Shannon index (H) for a given sample was calculated using the formula:

$$H = -\sum_{i=1}^{n} p_i \cdot \log(p_i) \tag{1}$$

where:

p_i : Proportion of reads for the ith MOTU cluster

n : Number of MOTUs

2.4. Generation of remote sensing variables

Several variables critical for fish assemblage in tropical rivers were calculated using remote sensing imagery. The river water surface temperature (RST) was determined using data from the thermal infrared sensor on Landsat-8 (Zong et al., 2023; Vanhellemont, 2020a). Subsequently, the Secchi depth (SD), serving as a proxy for water transparency, and the trophic state index (TSI), indicating nutrient enrichment, were calculated (Olmanson et al., 2008) utilizing both Level 1C (Top-of-Atmosphere Reflectance) and Level 2A (Surface Reflectance) products obtained from the Multispectral Instrument (MSI) aboard Sentinel-2, as well as data from the Landsat-8 Operational Land Imager (OLI) (Drusch et al., 2012; Page et al., 2019). Various vegetation indices were utilized to capture a comprehensive view of the vegetation surrounding the rivers. The forest canopy height was determined following the procedures outlined by Lang et al. (2023). The Normalized Difference Vegetation Index (NDVI) was calculated using data from the Landsat-8 OLI and the Enhanced Vegetation Index (EVI) (Sims et al., 2006) was extracted from MODIS products. In addition to these indices, total evapotranspiration and gross primary productivity (GPP) were retrieved from MODIS products. Evapotranspiration contributes to understanding the overall water and energy dynamics within the river-forest system (Kumagai et al., 2005) and GPP provides insights into ecosystem productivity and the efficiency of photosynthetic processes (Gebremichael and Barros, 2006). We calculated the slope of the water surface along the river channel, which can impact fish migration patterns and spawning behaviors (Jaramillo-Villa et al., 2010), as well as the elevation of the site. Further, the surrounding water area for each sample was computed by summing all water pixels within the predefined buffers. The analysis also incorporated the Human Modification Index (HMI) (Kennedy et al., 2019) which quantitatively measures human impact on the surrounding environment, encompassing various anthropogenic alterations. The variables were computed as the mean and standard deviation within a 3.5 km radius buffer zone around each sampling location, a range previously identified as useful in similar applications (Zong et al., 2023). Including standard deviation computations allows us to assess habitat heterogeneity as the standard deviation values capture each variable's degree of variability or dispersion within the specified radius. For water variables, non-water pixels were masked during the computation, ensuring a focused analysis of aquatic features. In contrast, water pixels were masked for terrestrial variables to concentrate on land-based characteristics. The temporal scope of the satellite imagery analysis extended over seven months for each river. This period covered three months before the initiation of sampling on a given river and three months afterward. The choice of this prolonged timeframe was imperative to secure cloud-free images, a particularly challenging task in tropical zones due to the persistent formation of robust cloud cover. All variables derived from time-sensitive satellite imagery were computed using data within this shared period to ensure temporal alignment, whereas variables based on static or pre-composited products were included as-is. The computations were executed using Google Earth Engine (Gorelick et al., 2017) (see Table 1).

2.5. Biodiversity variability

To discern whether variability in biodiversity metrics was primarily shaped by local or global factors, we first aggregated sample-based species richness and Shannon index values by river. The cross-sample median provides a central estimate of biodiversity within each river, while the spread indicates within-river heterogeneity. We then explored environmental gradients among sampling sites by performing a Principal Component Analysis (PCA) on the five most important predictors identified by the global models, using the `prcomp` function in R and retaining the first two principal components for visualization. We assessed the statistical significance of river-level separation in this

Table 1
Overview of the aquatic and terrestrial variables generated for the subsequent biodiversity analysis. Values were extracted as mean and standard deviation within a circular buffer of 3.5 km radius around the sample sites.

Abbreviation	Variable_Name	Description
RST/RST_std	Mean river water surface temperature/Standard deviation of river water surface temperature (Vanhellemont, 2020a)	Mean and standard deviation of river water surface temperature within the buffer
SD/SD_std	Mean Secchi depth/Standard deviation of Secchi depth (Page et al., 2019)	Mean and standard deviation of Secchi depth within the buffer
TSI/TSI_std	Mean trophic state index/Standard deviation of trophic state index (Page et al., 2019)	Mean and standard deviation of trophic state index within the buffer
NDVI/NDVI_std	Mean NDVI/Standard deviation of NDVI (Gessesse and Melesse, 2019)	Mean and standard deviation of normalized difference vegetation index within the buffer
EVI/EVI_std	Mean EVI/Standard deviation of EVI (Sims et al., 2006)	Mean and standard deviation of enhanced vegetation index within the buffer
GPP/GPP_std	Mean gross primary productivity/Standard deviation of GPP	Mean and standard deviation of gross primary productivity within the buffer
CH/CH_std	Mean canopy height/Standard deviation of canopy height (Lang et al., 2023)	Mean and standard deviation of canopy height within the buffer
EP/EP_std	Mean total evapotranspiration/Standard deviation of total evapotranspiration	Mean and standard deviation of total evapotranspiration within the buffer
Slope/Slope_std	Mean slope/Standard deviation of slope (Farr et al., 2007)	Mean and standard deviation of river surface slope within the buffer
Elev/Elev_std	Mean elevation/Standard deviation of elevation	Mean and standard deviation of elevation within the buffer
HMI/HMI_std	Mean human modification index/Standard deviation of human modification index (Kennedy et al., 2019)	Mean and standard deviation of human modification index within the buffer
WaterA	Sum of surrounding water area (Zong et al., 2023)	Total water area within the buffer

environmental space by applying permutational multivariate analysis of variance (PERMANOVA) with the `vegan::adonis2` function on the Euclidean distances of the scaled predictors. Finally, to investigate taxonomic composition, we created a family-level presence/absence matrix (1 if a family appeared in a sample, 0 otherwise), computed Jaccard dissimilarities with the `vegdist` function, and performed a Principal Coordinates Analysis (PCoA) using `cmdscale`, keeping the first two axes for plotting. Significance of river differences in community composition was similarly evaluated via PERMANOVA on the Jaccard distance matrix. We focused on the family level because the SWARM algorithm resulted in substantial gaps for finer taxonomic assignments, and 73% of samples had at least a family-level classification. MOTUs without an assigned family were excluded from the analysis.

2.6. *Model fitting*

To capture the non-linear relationships within ecological systems, we employed random forest regression models constructed using the R randomForest package, implementing Breiman's random forest algorithm (Breiman, 2001; Liaw and Wiener, 2002; R Core Team, 2021). Two distinct local models were developed for fish species richness and Shannon diversity, focusing on the Maroni and Oyapock rivers as representative cases due to their substantial sample sizes of 68 and 74, respectively. The other four rivers were excluded in this step due to their significantly smaller sample sizes, which were deemed insufficient for applying and testing meaningful random forest models in a river-specific context (Luan et al., 2020; van der Ploeg et al., 2014). Further, to assess the efficacy of model fitting across river systems on global scale, data from all six rivers was combined to create a comprehensive global model. All environmental variables were normalized within the datasets to address range differences. Subsequently, Pearson correlation coefficients were computed to assess potential correlations between the predictors. In instances where two variables exhibited a Pearson correlation exceeding 0.75, one of them was excluded based on domain knowledge. This aids in mitigating the impact of multicollinearity, thus preventing imprecision in subsequent analyses. To improve model efficiency and interpretability, recursive feature elimination (RFE) was implemented using the rfe function from the R caret package (Kuhn, 2008). The selection criterion was therefore based on minimizing the RMSE. The capacity of model fitting was then evaluated by using the whole dataset as training data. RMSE and percentual RMSE (RMSE normalized by the mean of the data) were computed to quantify the model fit. The coefficient of determination (R^2), representing the proportion of the variance in the dependent variable explained by the model, was employed to assess the explained variance in the training set (Chicco et al., 2021).

To uncover the main drivers shaping the patterns in fish species richness and Shannon index, the variable importance of the models were extracted directly from the randomForest object (Liaw and Wiener, 2002). The permutation-based MSE reduction was used as measure of importance (as implemented in the randomForest package). To further analyze the directional impact of the drivers, we generated partial dependency plots using the "partialPlot" function from the randomForest package (Liaw and Wiener, 2002). These plots illustrate the marginal effect of each predictor variable on the response while accounting for the average effects of all other variables in the model, and are based on the final selected models.

2.7. *Model evaluation*

2.7.1. *Evaluation on encountered rivers*

To assess the predictive performance of the models on unseen data, the dataset was initially divided randomly into five equally sized folds. For the global model, each fold contained proportional representation from individual rivers constituting the global dataset. Subsequently, distinct local and global models were iteratively trained on four folds and evaluated on the fifth fold, thus leading to a cross-validation procedure (Kohavi, 1995). The variables included in the models were those selected during the previous model fitting process using the combined dataset. To enhance robustness and account for variability in the predictions, bootstrap resampling was performed 10 times within each fold, generating multiple resampled datasets for training the models. 95% prediction intervals across the bootstrap iterations were computed for all samples. The performance was then evaluated using RMSE, percentual RMSE, and R^2 metrics. This methodology examines the models' capacity to accurately forecast outcomes on the rivers they were trained on, thus offering insights into their viability for generating biodiversity maps along rivers.

2.7.2. *Evaluation on unencountered rivers*

Additionally, to assess the feasibility of predicting biodiversity metrics in previously unsampled rivers, global models were consistently trained using data from five selected rivers, with each iteration excluding one river during the fitting process. Subsequently, the excluded

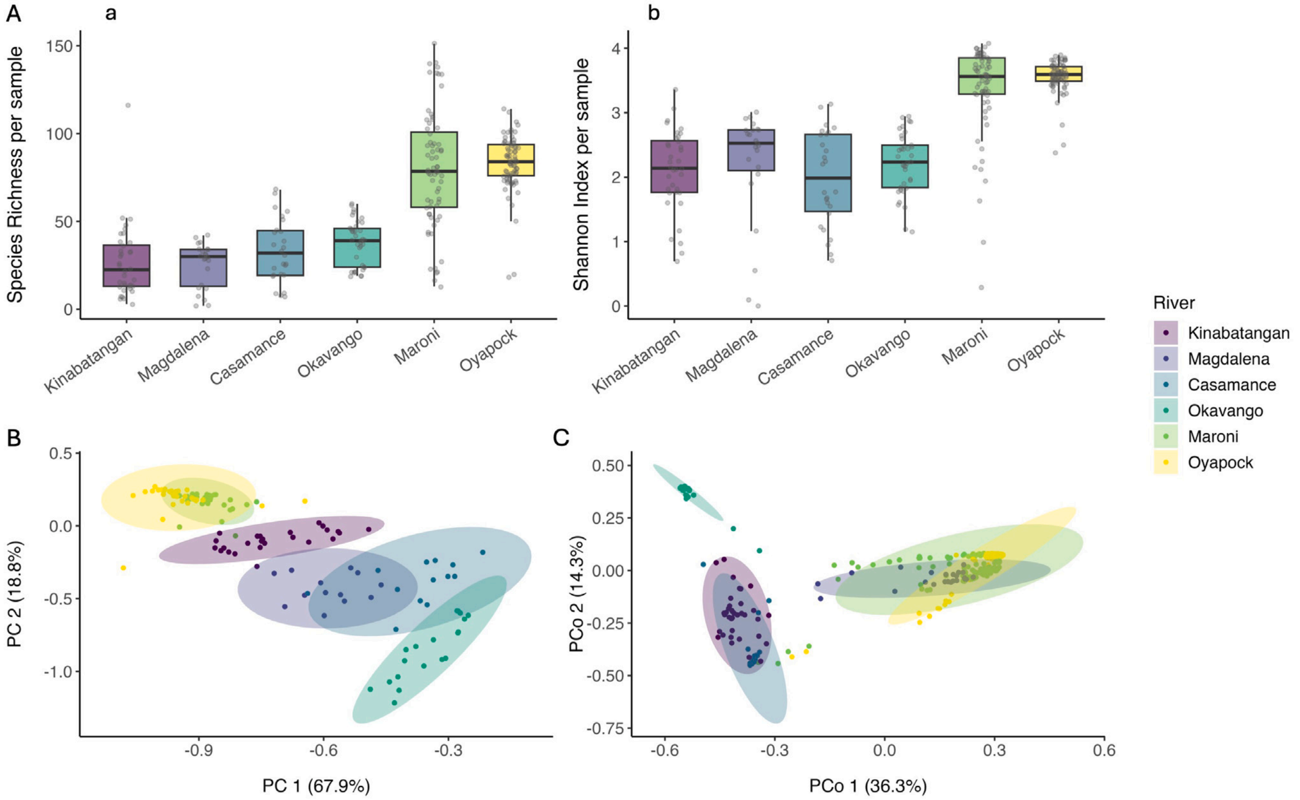

Fig. 2. (A) Boxplots of species richness (a) and Shannon index (b) per sample across each river in the dataset. Boxes span the interquartile range, the horizontal line marks the median, whiskers extend to 1.5 times the interquartile range, and points represent individual samples. (B) Principal Component Analysis (PCA) of the environmental variables at each sampling site, focusing on the five most important predictors identified by the global richness models (canopy height, elevation, human modification index, river surface temperature, and the standard deviation of evapotranspiration). The first two components are displayed, with the percentage of variance explained in parentheses. Each point represents a sample colored by river, and 95% confidence ellipses highlight group clustering. (C) Principal Coordinates Analysis (PCoA) based on Jaccard dissimilarity of family-level taxonomic composition across samples. The first two axes are shown, with variance explained in parentheses. Each point corresponds to a sample colored by river, and 95% confidence ellipses illustrate clustering patterns among rivers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

river was used as a testing set for predicting the biodiversity metrics. To quantify uncertainty, bootstrap resampling was performed 10 times for the training data in each iteration. This approach provides insights into the models' predictive performance on rivers not included in the training phase and their potential application in generating biodiversity assessments across tropical river ecosystems.

3. Results

3.1. Biodiversity variability

The Oyapock and Maroni rivers exhibit the highest median species richness, with 84 and 78.5 species per sample, respectively, while the Kinabatangan shows the lowest at 22.5, highlighting significant variation across the studied rivers (Fig. 2A,a). The Maroni River, in particular, displays considerable variability, with species counts ranging from as few as 13 to as many as 151 in individual samples (Fig. 2A,a). For the Shannon index per sample, the Oyapock and Maroni rivers again stand out, recording the highest median values of 3.59 and 3.56, respectively (Fig. 2A,b). Environmental conditions among sampling sites vary, with notable clustering observed between the Maroni and Oyapock, as well as between the Magdalena and Casamance rivers (Fig. 2B). At the family level, the South American rivers — Maroni, Oyapock, and Magdalena — exhibit high taxonomic similarity. Interestingly, the Kinabatangan and Casamance rivers also show similarity in their family-level composition, whereas the Okavango remains distinct in PcOA space. PERMANOVA confirmed that river-level separations were statistically significant in both the environmental (PCA) and community (PCoA) analyses ($p < 0.001$).

3.2. Model performance in capturing fish diversity patterns

The species richness and Shannon index models demonstrate good fit to the data across diverse contexts (Fig. 4). The global models exhibit an explained variance (R^2) of 95% for richness and 90% for the Shannon model, complemented by low RMSE values (Fig. 4a,b), which is strong considering the extensive distribution of rivers across three continents and the consequently relatively limited sample size. The local models applied to the Maroni and Oyapock rivers score explained variances of 91% and 81% in richness models and 74% and 82% in Shannon index models, respectively (Fig. 4c–f). The Maroni models reveal a significantly wider distribution of values compared to the Oyapock models, underscoring the pronounced ecological gradient within the Maroni river (Fig. 2).

Canopy height stands out as the predominant predictor for both global species richness and the Shannon index, with higher values positively affecting both metrics. The Human Modification Index and river surface temperature also have high predictive power in the models of both metrics, both showing a negative effect. Additionally, global richness is strongly influenced by elevation and the standard deviation in total evapotranspiration whilst in the Shannon index model, the total sum of surrounding water area and the standard deviation in canopy height further emerge as crucial (see Fig. 3).

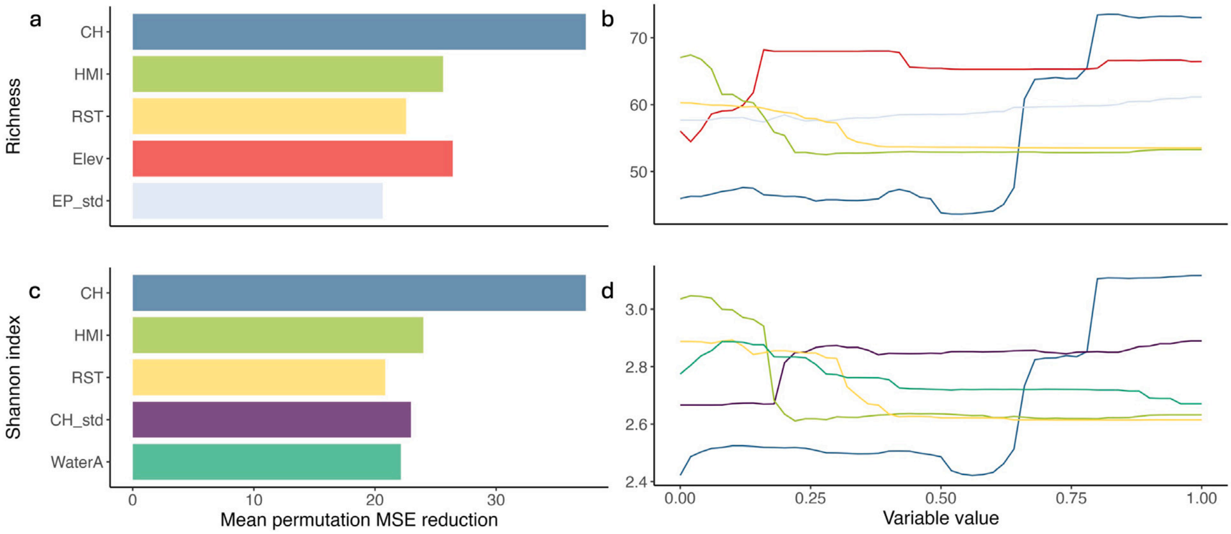

Fig. 3. Variable importance via permutation-based mean squared error (MSE) reduction for the top five variables in the global models of species richness (a) and Shannon index (c), with corresponding partial dependency plots (b,d) (Wickham et al., 2019). Equivalent plots for the local models are provided in Appendix A.2.

3.3. Predictive performance of integrated models

3.3.1. Evaluation on encountered rivers

In the five-fold cross-validation testing, the global models demonstrated good performance, achieving an explained variance of 0.82 ± 0.06 for the species richness model and 0.71 ± 0.12 for the Shannon index model (Fig. 4a–b). On a more localized scale, the Maroni models exhibited efficacy in predicting the richness of the reserved samples, achieving an R^2 value of 0.73 ± 0.11. The Shannon index prediction is less successful with 0.37 ± 0.23 (Fig. 4c–d). The Oyapock models displayed a slight decline in R^2, registering values of 0.19 ± 0.39 and 0.12 ± 0.54, respectively (Fig. 4e–f). However, they showcased a lower percentage RMSE than the corresponding global and Maroni river models. This discrepancy can be attributed to the observably lower overall variance in the Oyapock eDNA data.

3.3.2. Evaluation on unencountered rivers

When predicting fish species richness and Shannon index values for rivers excluded from the training process, individual models show limited accuracy (Fig. 5). The negative individual R^2 values indicate performance below a baseline model that predicts the mean. However, a broader view reveals a noticeable trend: collectively, the models generate predictions that fall within the expected value range (Fig. 5). Specifically, the combined species richness predictions yield an R^2 value of 0.39, indicating moderate predictive power. Similarly, the aggregate Shannon index predictions result in an R^2 of 0.46. These results indicate that while individual models may struggle with accuracy, their combined predictions offer an approximation of the biodiversity metric values in unencountered rivers.

4. Discussion

This study demonstrates the efficacy of combining eDNA metabarcoding data with environmental variables derived from remote sensing for capturing and predicting fish biodiversity patterns in tropical river ecosystems. The integrated models demonstrate a good fit with their training data on both local and global scales. Additionally, particularly the richness models are effective in predicting fish biodiversity within the rivers on which they are trained. This offers potential for generating continuous biodiversity maps which can guide conservation strategies in those highly pressured ecosystems (see Fig. 6). Furthermore, while the global models currently fall short of achieving accurate predictions for unencountered rivers, they exhibit the capacity to forecast large-scale patterns and trends. This suggests the considerable potential of the approach, especially when extended with additional samples becoming available.

4.1. Biodiversity variability

The results reveal variation in both median fish species richness and Shannon index values across the studied rivers, indicative of strong cross-river differences in baseline biodiversity. One factor contributing to those differences is the variability in environmental conditions across the rivers (Fig. 2B). For instance, the Maroni and Oyapock rivers traverse dense rainforest, while the Casamance river flows through mangrove forests and grasslands. This environmental variability influences fish habitats; for example, riparian vegetation provides fruits, leaves, and insects directly consumed by fish or integrated into the river food web via macroinvertebrate shredders, which are in turn crucial food sources for fish (Vannote et al., 1980). Beyond the environment, biogeographical factors cause differences in river baseline biodiversity as the separation of river basins by impassable land and oceanic barriers leads to unique extinction and speciation processes within each insular river system (Oberdorff et al., 2011; Dias et al., 2014). This is also reflected in the variation in taxonomic composition across the rivers (Fig. 2C.) While cross-river differences are substantial, variability persists within rivers due to environmental conditions differing across sampled locations (Fig. 2B). The significant variation in baseline biodiversity among rivers underscores the necessity of recognizing ecological dynamics not only at local but also at global scales for informative model development. While the variables employed in this study adequately account for local conditions influencing fish habitats by integrating various factors in the aquatic and terrestrial surroundings of the sample sites (Zong et al., 2023), future research could enhance consideration of global determinants by incorporating biogeographic factors, which are well-documented for their strong influence on large-scale biodiversity patterns (Oberdorff et al., 2011; Collen et al., 2014).

4.2. Model performance in capturing fish diversity patterns

The effectiveness of remote sensing-derived environmental variables in capturing the distribution of biodiversity in tropical rivers,

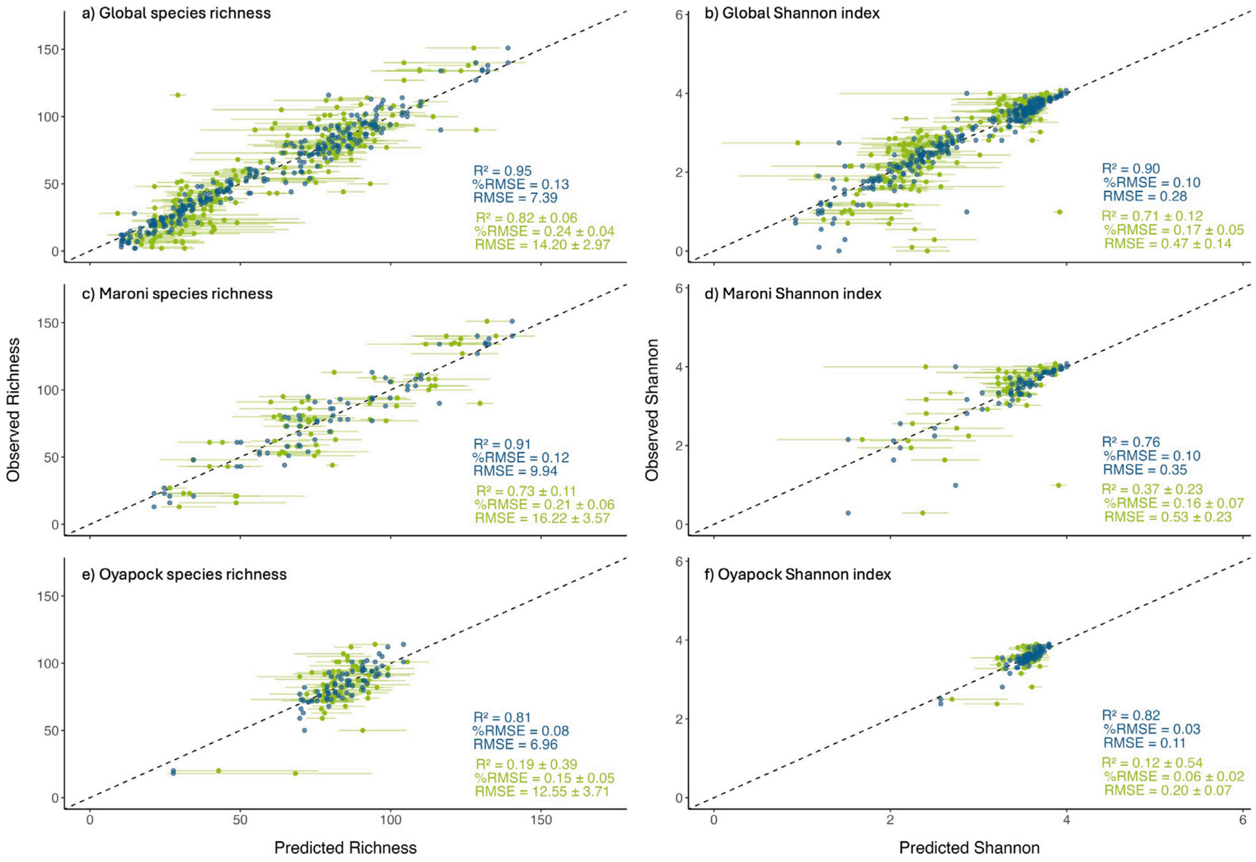

Fig. 4. Combined visual representation of model fit (blue) and five-fold cross-validation predictions (green) for species richness (left) and Shannon diversity index (right). The black dashed line represents the identity line, indicating perfect agreement between predicted and observed values. For cross-validation predictions, points represent the mean of 10 bootstrap resampling iterations, with 95% prediction intervals shown as bars for each sample. Evaluation metrics (R^2, RMSE, and %RMSE) for cross-validation represent the mean and standard deviation across the five folds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

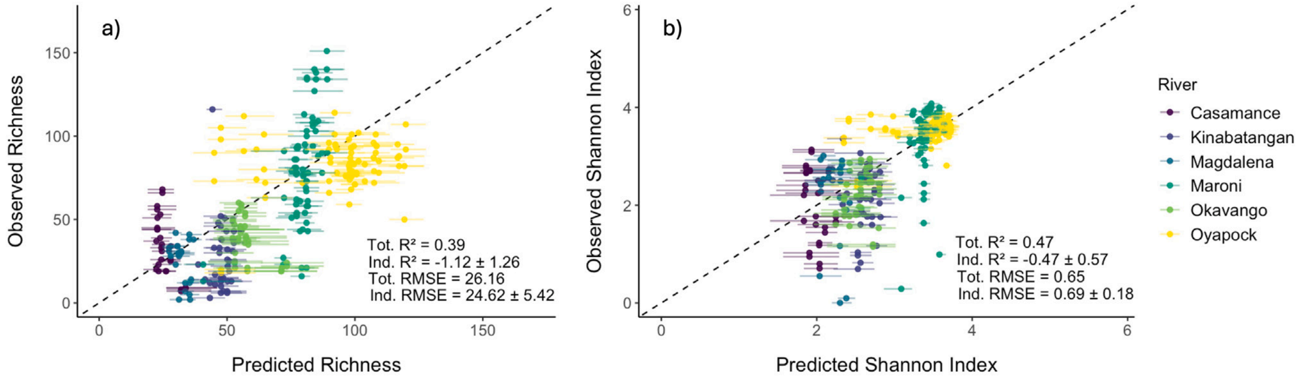

Fig. 5. Combined predictions for species richness (a) and Shannon index (b) models, trained on data from five rivers and applied to predict values for sites in the sixth river. The black dashed line represents the identity line, indicating perfect agreement between predicted and observed values. Points represent the mean of 10 bootstrap resampling iterations, with 95% prediction intervals shown as error bars for each sample. Performance metrics are reported as the mean and standard deviation across the individual rivers, as well as overall metrics for the combined predictions. These results provide an overview of model performance on previously unencountered rivers.

particularly regarding species richness, is evident from the robust fit of the developed models to the training data. The strongly positive impact of canopy height can be attributed to its role in providing shade, food, and thermal buffering to the river, as well as influencing nutrient dynamics and water quality, all of which are crucial for fish species assemblage (Beltrão et al., 2009; Tabacchi et al., 2000; Vannote et al., 1980). Conversely, the Human Modification Index negatively impacts biodiversity, as landscape changes from human activities like deforestation and pollution disrupt tropical river systems and their fish communities (Arantes et al., 2018; Leitão et al., 2018). The sharp decline in the initial section of the response curve indicates that pristine river ecosystems are particularly favorable for high fish biodiversity, and even minor increases in human impact can result in a severe loss of biodiversity. Elevated river surface temperatures might negatively impact fish richness and the Shannon index by disrupting the thermal regimes critical for many fish species, particularly in tropical regions where these species often have narrow thermal niches and are constantly close to their upper thermal limits (Lapointe et al., 2018; Stevens, 1989). Elevation emerges as an important predictor in the

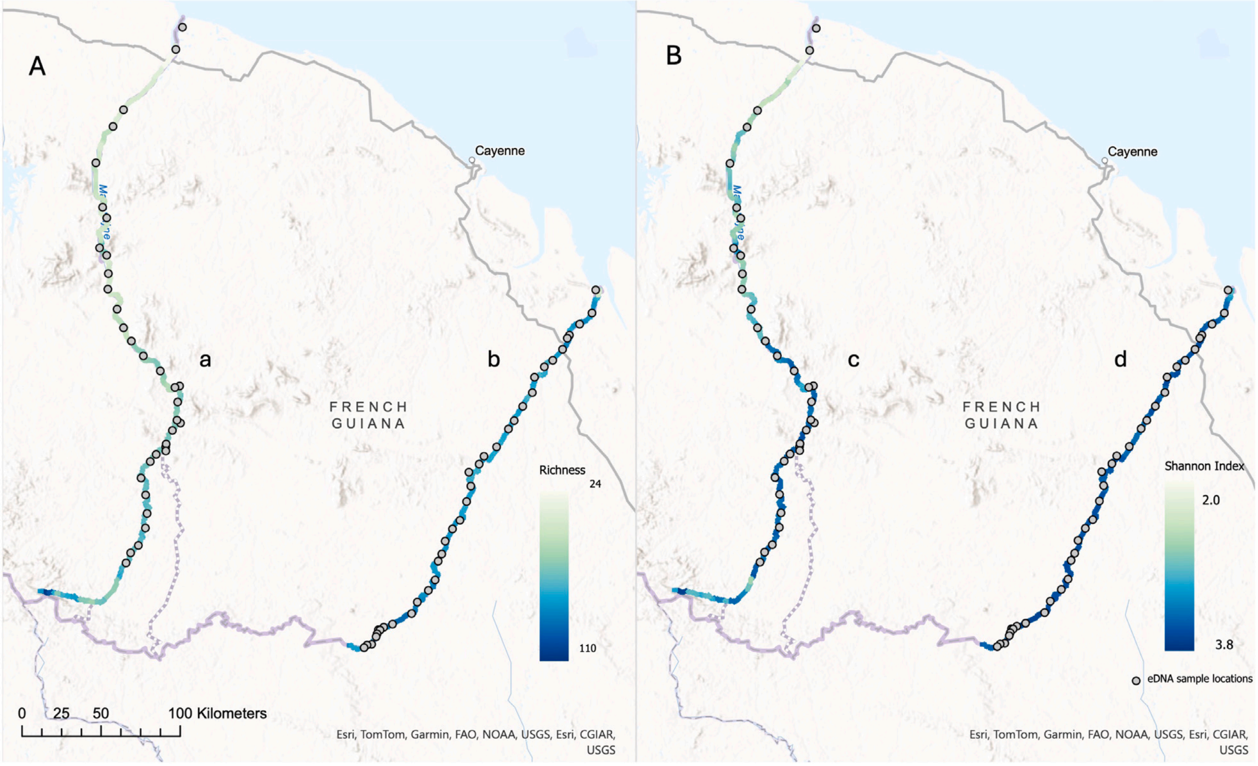

Fig. 6. Spatial representation of model predictions for species richness (A) and Shannon index (B) in the Maroni (a, c) and Oyapock (b, d) rivers in French Guiana. Predictions were generated using local models trained on each river separately.

global species richness model, not necessarily because it directly impacts fish occurrence, but possibly because it explains baseline richness differences between rivers (Fig. 2).

4.3. Predictive performance of integrated models

The richness models effectively predict rivers on which they are trained. This presents a promising avenue for assessing biodiversity patterns in those ecosystems and generating insights valuable for facilitating effective conservation strategies. The RMSE values of the global, Maroni, and Oyapock species richness models are comparable, suggesting inherent challenges, such as biases and uncertainties associated with using eDNA to capture community compositions (Beng and Corlett, 2020). A similar pattern is observed for the Shannon index models, but here, the RMSE is significantly larger for the global model compared to the local ones. This outcome is anticipated, given that the Shannon index is likely more susceptible to biases introduced during DNA extraction, PCR amplification, or variations in total read numbers across different rivers (Beng and Corlett, 2020).

While models perform well on rivers encountered during training, predicting outcomes for geographically new regions presents a greater challenge. The models struggle to accurately predict outcomes for rivers not encountered during training, however, the forecasts generally align with the anticipated range (Fig. 5).. The constrained transferability observed in this context is expected, given the considerable geographical distances and distinct tropical biome types among the studied rivers (Fig. 1). Future models could address the challenge of generalizing across rivers by incorporating biome types or geographic regions as categorical predictor variables. These variables could serve as proxies for fundamental biodiversity baselines that vary across regions. However, such an approach is heavily dependent on data availability, as categorical predictors require adequate representation during model training. In our study, most biome types were represented by only one river, and at a continental scale, we had data for only a single river from Asia. This lack of representation made it impossible to evaluate model performance on rivers outside the training set without including them in the training data. As the use of eDNA sampling expands (Beng and Corlett, 2020; Ogden, 2021) and river eDNA datasets get larger, future models can be enabled to include biome or regional predictors. With such refinements, future models might go beyond predicting outcomes within individual rivers but also achieve robust generalization across diverse river systems, offering insights into biodiversity patterns in under-sampled regions. As remote sensing data is regularly updated (Ozesmi and Bauer, 2002; Zong et al., 2023), the integration of remote sensing with expanded eDNA datasets has the potential to advance riverine ecosystem monitoring, reducing the need for repeated on-site sampling. However, achieving these goals will require further refinement of modeling approaches and the accumulation of larger and more representative datasets.

4.4. Limitations

We utilized the SWARM algorithm to mitigate the limited depth of reference databases for tropical fish species. However, SWARM has inherent limitations, as some MOTU clusters may not correspond to distinct biological species and can occasionally result from PCR and sequencing errors (Polanco Fernández et al., 2021). Therefore, for accurate species-level identification, developing comprehensive reference databases for tropical environments remains essential. Additionally, our sampling effort was more extensive on the Maroni and Oyapock rivers compared to other rivers, biasing our dataset toward these two systems. Further, exact temporal alignment of environmental variable extraction was not possible due to cloud cover. However, extracting values within a six-month window centered on the sampling month should capture the broader environmental conditions during sampling.

5. Conclusion and outlook

Our study demonstrates that combining eDNA and remote sensing data offers a promising framework for predicting fish biodiversity in tropical rivers. The integrated approach effectively estimates key biodiversity metrics — particularly species richness — in sampled rivers, making it a valuable tool for biodiversity mapping. While predictions for unsampled rivers currently show limited precision, they capture broad-scale biodiversity patterns. As eDNA data from a wider range of tropical rivers become available, model accuracy will likely improve, supporting more reliable forecasts even in previously unsampled systems. Furthermore, our findings indicate that high-canopy habitats and limited human disturbance sustain more diverse fish communities, underscoring the importance of preserving near-pristine conditions in tropical rivers and their surrounding landscapes.

CRediT authorship contribution statement

Robin Bauknecht: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Loïc Pellissier:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Sébastien Brosse:** Writing – review & editing, Resources. **Vincent Prié:** Writing – review & editing, Resources. **Manuel Lopes-Lima:** Writing – review & editing, Resources. **Pedro Beja:** Writing – review & editing, Resources. **Monika K. Goralczyk:** Writing – review & editing, Resources. **Andrea Polanco Fernandez:** Writing – review & editing, Resources. **Jorge A. Moreno-Tilano:** Writing – review & editing, Resources. **Rafik Neme:** Writing – review & editing, Resources. **Mailyn A. Gonzalez:** Writing – review & editing, Resources. **Shuo Zong:** Writing – review & editing, Supervision, Resources, Methodology, Visualization, Conceptualization.

Declaration of competing interest

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

Acknowledgments

We thank Eric Bairrão Ruivo and Gabriel Michelin for providing the eDNA data from their sampling campaigns on the Casamance and Kinabatangan rivers. We also thank everyone who contributed to the fieldwork and laboratory analysis. Additionally, we appreciate the open data policies of the ESA Copernicus Programme, the USGS Landsat mission, and the NASA MODIS project, which were instrumental to our research.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103251.

Data availability

The code and data used for this analysis is publicly available on Zenodo: https://doi.org/10.5281/zenodo.15869405

.

References

Altermatt, F., Little, C.J., Mächler, E., Wang, S., Zhang, X., Blackman, R.C., 2020. Uncovering the complete biodiversity structure in spatial networks: the example of riverine systems. Oikos 129, 607–618, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.06806.

Arantes, C.C., Winemiller, K.O., Petrere, M., Castello, L., Hess, L.L., Freitas, C.E.C., 2018. Relationships between forest cover and fish diversity in the amazon river floodplain. J. Appl. Ecol. 55, 386–395, URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.12967.

Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation dna sequencing. Mol. Ecol. 21, 2039–2044, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2012.05519.x.

Beltrão, G., Medeiros, E., Ramos, R., 2009. Effects of riparian vegetation on the structure of the marginal aquatic habitat and the associated fish assemblage in a tropical brazilian reservoir. Biota Neotrop. 9, 37–43.

Beng, K.C., Corlett, R.T., 2020. Applications of environmental dna (edna) in ecology and conservation: opportunities, challenges and prospects. Biodivers. Conserv. 29, 2089–2121.

Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., Radeloff, V.C., 2009. Remote sensing of vegetation 3-d structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. J. Geophys. Res.: Biogeosciences 114.

Boivin-Delisle, D., Laporte, M., Burton, F., Dion, R., Normandeau, E., Bernatchez, L., 2021. Using environmental dna for biomonitoring of freshwater fish communities: Comparison with established gillnet surveys in a boreal hydroelectric impoundment. Environ. DNA 3, 105–120, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.135.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. Obitools: a unix-inspired software package for dna metabarcoding. Mol. Ecol. Resour. 16, 176–182, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12428.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Campbell, G., Phinn, S.R., Dekker, A.G., Brando, V.E., 2011. Remote sensing of water quality in an australian tropical freshwater impoundment using matrix inversion and meris images. Remote Sens. Environ. 115, 2402–2414.

Cantera, I., Coutant, O., Jézéquel, C., Decotte, J.-B., Dejean, T., Iribar, A., Vigouroux, R., Valentini, A., Murienne, J., Brosse, S., 2022. Low level of anthropization linked to harsh vertebrate biodiversity declines in Amazonia. Nat. Commun. 13, 3290. http://dx.doi.org/10.1038/s41467-022-30842-2.

Carrizo, S.F., Jähnig, S.C., Bremerich, V., Freyhof, J., Harrison, I., He, F., Langhans, S.D., Tockner, K., Zarfl, C., Darwall, W., 2017. Freshwater megafauna: Flagships for freshwater biodiversity under threat. BioScience 67, 919–927.

Carvajal-Quintero, J.D., Escobar, F., Alvarado, F., Villa-Navarro, F.A., Jaramillo-Villa, Ú., Maldonado-Ocampo, J.A., 2015. Variation in freshwater fish assemblages along a regional elevation gradient in the northern andes, colombia. Ecol. Evol. 5, 2608–2620, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.1539.

Cetra, M., Petrere, Jr., M., 2007. Associations between fish assemblage and riparian vegetation in the corumbataíriver basin (sp). Braz. J. Biol. 67, 191–195.

Cheng, R., Luo, Y., Li, Q., Zhang, Y., Liu, Z., Chen, Q., Li, Y., Shen, Y., 2023. Application of edna metabarcoding for monitoring the fish diversity of the jiangjin to fuling section of the upper reaches of the yangtze river. Hydrobiologia 850, 4067–4088.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. PeerJ Comput. Sci. 7, e623.

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., Iribar, A., Taberlet, P., Vigouroux, R., Brosse, S., 2019. Unlocking biodiversity and conservation studies in high-diversity environments using environmental dna (edna): A test with guianese freshwater fishes. Mol. Ecol. Resour. 19, 27–46.

Collen, B., Whitton, F., Dyer, E.E., Baillie, J.E.M., Cumberlidge, N., Darwall, W.R.T., Pollock, C., Richman, N.I., Soulsby, A.-M., Böhm, M., 2014. Global patterns of freshwater species diversity, threat and endemism. Glob. Ecol. Biogeogr. 23, 40–51, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12096.

Coutant, O., Jézéquel, C., Mokany, K., Cantera, I., Covain, R., Valentini, A., Dejean, T., Brosse, S., Murienne, J., 2023. Environmental dna reveals a mismatch between diversity facets of amazonian fishes in response to contrasting geographical, environmental and anthropogenic effects. Global Change Biol. 29, 1741–1758, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.16533.

Crippen, R.E., 1990. Calculating the vegetation index faster. Remote Sens. Environ. 34, 71–73, URL: https://www.sciencedirect.com/science/article/pii/003442579090085Z.

Darwall, W., Bremerich, V., De Wever, A., Dell, A.I., Freyhof, J., Gessner, M.O., Grossart, H.-P., Harrison, I., Irvine, K., Jähnig, S.C., Jeschke, J.M., Lee, J.J., Lu, C., Lewandowska, A.M., Monaghan, M.T., Nejstgaard, J.C., Patricio, H., Schmidt-Kloiber, A., Stuart, S.N., Thieme, M., Tockner, K., Turak, E., Weyl, O., 2018. The alliance for freshwater life: A global call to unite efforts for freshwater biodiversity science and conservation. Aquat. Conserv.: Mar. Freshw. Ecosyst. 28, 1015–1022, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aqc.2958.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental dna metabarcoding: Transforming how we survey animal and plant communities. Mol. Ecol. 26, 5872–5895, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.14350.

Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.-C., Altermatt, F., 2016. Environmental dna reveals that rivers are conveyer belts of biodiversity information. Nat. Commun. 7 (12544).

Dias, M.S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J.-F., Brosse, S., Grenouillet, G., Tedesco, P.A., 2014. Global imprint of historical connectivity on freshwater fish biodiversity. Ecol. Lett. 17, 1130–1140, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12319.

Domisch, S., Amatulli, G., Jetz, W., 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Sci. Data 2, 150073.

Drusch, M., Bello, U.D., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. Remote Sens. Environ. 120, 25–36.

Dubey, V.K., Sarkar, U.K., Pandey, A., Sani, R., Lakra, W.S., 2012. The influence of habitat on the spatial variation in fish assemblage composition in an unimpacted tropical river of ganga basin, india. Aquat. Ecol. 46, 165–174.

Dudgeon, D., 2000. The ecology of tropical asian rivers and streams in relation to biodiversity conservation. Annu. Rev. Ecol. Syst. 31, 239–263.

Encalada, A.C., Flecker, A.S., Poff, N.L., Suárez, E., Herrera-R, G.A., Ríos-Touma, B., Jumani, S., Larson, E.I., Anderson, E.P., 2019. A global perspective on tropical montane rivers. Science 365, 1124–1129.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. Rev. Geophys. 45.

Ferguson, G.J., Ward, T.M., Ye, Q., Geddes, M.C., Gillanders, B.M., 2013. Impacts of drought, flow regime, and fishing on the fish assemblage in southern australia's largest temperate estuary. Estuaries Coasts 36, 737–753.

Frøslev, T.G., Kjøller, R., Bruun, H.H., Ejrnæs, R., Brunbjerg, A.K., Pietroni, C., Hansen, A.J., 2017. Algorithm for post-clustering curation of dna amplicon data yields reliable biodiversity estimates. Nat. Commun. 8 (1188).

Gebremichael, M., Barros, A., 2006. Evaluation of modis gross primary productivity (gpp) in tropical monsoon regions. Remote Sens. Environ. 100, 150–166.

Gessesse, A.A., Melesse, A.M., 2019. Chapter 8 - temporal relationships between time series chirps-rainfall estimation and emodis-ndvi satellite images in amhara region, ethiopia. In: Melesse, A.M., Abtew, W., Senay, G. (Eds.), Extreme Hydrology and Climate Variability. Elsevier, pp. 81–92, URL: https://www.sciencedirect.com/science/article/pii/B9780128159989000087.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Hugue, F., Lapointe, M., Eaton, B., Lepoutre, A., 2016. Satellite-based remote sensing of running water habitats at large riverscape scales: Tools to analyze habitat heterogeneity for river ecosystem management. Geomorphology 253, 353–369, URL: https://www.sciencedirect.com/science/article/pii/S0169555X15301902.

Humpl, M., Pivnička, K., 2006. Fish assemblages as influenced by environmental factors in streams in protected areas of the czech republic. Ecol. Freshw. Fish 15, 96–103, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0633.2006.00126.x.

Jaramillo-Villa, U., Maldonado-Ocampo, J.A., Escobar, F., 2010. Altitudinal variation in fish assemblage diversity in streams of the central andes of colombia. J. Fish Biol. 76, 2401–2417, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.2010.02629.x.

Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Baruch-Mordo, S., Kiesecker, J., 2019. Managing the middle: A shift in conservation priorities based on the global human modification gradient. Global Change Biol. 25, 811–826.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2. IJCAI'95, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, pp. 1137–1143.

Kuhn, M., 2008. Building predictive models in r using the caret package. J. Stat. Softw. 28, 1–26, URL: https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Kumagai, T., Saitoh, T.M., Sato, Y., Takahashi, H., Manfroi, O.J., Morooka, T., Kuraji, K., Suzuki, M., Yasunari, T., Komatsu, H., 2005. Annual water balance and seasonality of evapotranspiration in a bornean tropical rainforest. Agricult. Forest. Meterol. 128, 81–92.

Lacoursière-Roussel, A., Côté, G., Leclerc, V., Bernatchez, L., 2016. Quantifying relative fish abundance with edna: a promising tool for fisheries management. J. Appl. Ecol. 53, 1148–1157, URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.12598.

Lamb, E.G., Bayne, E., Holloway, G., Schieck, J., Boutin, S., Herbers, J., Haughland, D.L., 2009. Indices for monitoring biodiversity change: Are some more effective than others? Ecol. Indic. 9, 432–444.

Lang, N., Jetz, W., Schindler, K., Wegner, J.D., 2023. A high-resolution canopy height model of the earth. Nat. Ecol. Evol. 7, 1778–1789.

Lapointe, D., Cooperman, M.S., Chapman, L.J., Clark, T.D., Val, A.L., Ferreira, M.S., Balirwa, J.S., Mbabazi, D., Mwanja, M., Chhom, L., Hannah, L., Kaufman, L., Farrell, A.P., Cooke, S.J., 2018. Predicted impacts of climate warming on aerobic performance and upper thermal tolerance of six tropical freshwater fishes spanning three continents. Conserv. Physiol. 6, coy056, URL: https://doi.org/10.1093/conphys/coy056.

Lázaro-Vázquez, A., Castillo, M.M., Jarquín-Sánchez, A., Carrillo, L., Capps, K.A., 2018. Temporal changes in the hydrology and nutrient concentrations of a large tropical river: Anthropogenic influence in the lower grijalva river, mexico. River Res. Appl. 34, 649–660, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.3301.

Leitão, R.P., Zuanon, J., Mouillot, D., Leal, C.G., Hughes, R.M., Kaufmann, P.R., Villéger, S., Pompeu, P.S., Kasper, D., de Paula, F.R., Ferraz, S.F.B., Gardner, T.A., 2018. Disentangling the pathways of land use impacts on the functional structure of fish assemblages in amazon streams. Ecography 41, 219–232.

Li, J., Huang, L., Zou, L., Kano, Y., Sato, T., Yahara, T., 2012. Spatial and temporal variation of fish assemblages and their associations to habitat variables in a mountain stream of north tiaoxi river, china. Environ. Biol. Fishes 93, 403–417.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2, 18–22, URL: https://CRAN.R-project.org/doc/Rnews/.

Luan, J., Zhang, C., Xu, B., Xue, Y., Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. Fish. Res. 227, 105534, URL: https://www.sciencedirect.com/science/article/pii/S0165783620300515.

Lyet, A., Pellissier, L., Valentini, A., Dejean, T., Hehmeyer, A., Naidoo, R., 2021. Edna sampled from stream networks correlates with camera trap detection rates of terrestrial mammals. Sci. Rep. 11 (11362).

Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2014. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2, e593.

Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K., Cook, M., Schott, J.R., 2018. An operational land surface temperature product for landsat thermal data: Methodology and validation. IEEE Trans. Geosci. Remote Sens. 56, 5717–5735.

Margalef, R., 1958. Information theory in ecology. Gen. Syst. 3, 36–71.

Marques, V., Guérin, P., Rocle, M., Valentini, A., Manel, S., Mouillot, D., Dejean, T., 2020. Blind assessment of vertebrate taxonomic diversity across spatial scales by clustering environmental dna metabarcoding sequences. Ecography 43, 1779–1790.

Marques, V., Milhau, T., Albouy, C., Dejean, T., Manel, S., Mouillot, D., Juhel, J., 2021. Gapedna: Assessing and mapping global species gaps in genetic databases for edna metabarcoding. Diversity and Distributions 27, 1880–1892.

Marshall, S., Elliott, M., 1998. Environmental influences on the fish assemblage of the humber estuary, u.k. Estuar. Coast. Shelf Sci. 46, 175–184.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 17 (10).

Mathon, L., Marques, V., Manel, S., Albouy, C., Andrello, M., Boulanger, E., Deter, J., Hocdé, R., Leprieur, F., Letessier, T.B., Loiseau, N., Maire, E., Valentini, A., Vigliola, L., Baletaud, F., Bessudo, S., Dejean, T., Faure, N., Guerin, P.-E., Jucker, M., Juhel, J.-B., Kadarusman, Polanco F., A., Pouyaud, L., Schwörer, D., Thompson, K.F., Troussellier, M., Sugeha, H.Y., Velez, L., Zhang, X., Zhong, W., Pellissier, L., Mouillot, D., 2023. The distribution of coastal fish edna sequences in the anthropocene. Glob. Ecol. Biogeogr. 32, 1336–1352, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13698.

McIntyre, P.B., Liermann, C.A.R., Revenga, C., 2016. Linking freshwater fishery management to global food security and biodiversity conservation. Proc. Natl. Acad. Sci. 113, 12880–12885.

Mertes, L.A., Daniel, D.L., Melack, J.M., Nelson, B., Martinelli, L.A., Forsberg, B.R., 1995. Spatial patterns of hydrology, geomorphology, and vegetation on the floodplain of the amazon river in brazil from a remote sensing perspective. Geomorphology 13, 215–232, URL: https://www.sciencedirect.com/science/article/pii/0169555X95000387. Biogeomorphology, Terrestrial and Freshwater Systems.

Miqueleiz, I., Ariño, A.H., Miranda, R., 2023. Spatial priorities for freshwater fish conservation in relation to protected areas. Aquat. Conserv.: Mar. Freshw. Ecosyst. 33, 1028–1038.

Montag, L.F.A., Winemiller, K.O., Keppeler, F.W., Leão, H., Benone, N.L., Torres, N.R., Prudente, B.S., Begot, T.O., Bower, L.M., Saenz, D.E., Lopez-Delgado, E.O., Quintana, Y., Hoeinghaus, D.J., Juen, L., 2019. Land cover, riparian zones and instream habitat influence stream fish assemblages in the eastern amazon. Ecol. Freshw. Fish 28, 317–329.

Ngoye, E., Machiwa, J.F., 2004. The influence of land-use patterns in the ruvu river watershed on water quality in the river system. Phys. Chem. Earth Parts A/B/C 29, 1161–1166.

Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., Dürr, H.H., 2011. Global and regional patterns in riverine fish species richness: A review. Int. J. Ecol. 2011, 1–12.

Offem, B.O., Samsons, Y.A., Omoniyi, I.T., 2009. Fish composition and abundance in the wetlands of cross river, nigeria. Aquat. Ecol. 43, 1155–1166.

Ogden, L.E., 2021. The Emergence of eDNA: An interdisciplinary tool helps monitor biodiversity and health. BioScience 72, 5–12, URL: https://doi.org/10.1093/biosci/biab120.

Olmanson, L.G., Bauer, M.E., Brezonik, P.L., 2008. A 20-year landsat water clarity census of minnesota's 10 000 lakes. Remote Sens. Environ. 112, 4086–4097.

Ozesmi, S.L., Bauer, M.E., 2002. Satellite remote sensing of wetlands. Wetl. Ecol. Manag. 10, 381–402, URL: https://doi.org/10.1023/A:1020908432489.

Page, B.P., Olmanson, L.G., Mishra, D.R., 2019. A harmonized image processing workflow using sentinel-2/msi and landsat-8/oli for mapping water clarity in optically variable lake systems. Remote Sens. Environ. 231, 111284.

Pawlowski, J., Apothéloz-Perret-Gentil, L., Altermatt, F., 2020. Environmental dna: What's behind the term? clarifying the terminology and recommendations for its future use in biomonitoring. Mol. Ecol. 29, 4258–4264.

Pawlowski, J., Bonin, A., Boyer, F., Cordier, T., Taberlet, P., 2021. Environmental dna for biomonitoring. Mol. Ecol. 30, 2931–2936.

Picado Barboza, J., Umaña Villalobos, G., 2018. Fish assemblages and their ecological traits along an elevational gradient in the Río Pacuare, Costa Rica. Rev. Biol. Trop. 66, S132–S152, URL: https://revistas.ucr.ac.cr/index.php/rbt/article/view/33269.

van der Ploeg, T., Austin, P.C., Steyerberg, E.W., 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med. Res. Methodol. 14 (137).

Polanco Fernández, A., Marques, V., Fopp, F., Juhel, J.-B., Borrero-Pérez, G.H., Cheutin, M.-C., Dejean, T., González Corredor, J.D., Acosta-Chaparro, A., Hocdé, R., Eme, D., Maire, E., Spescha, M., Valentini, A., Manel, S., Mouillot, D., Albouy, C., Pellissier, L., 2021. Comparing environmental dna metabarcoding and underwater visual census to monitor tropical reef fishes. Environ. DNA 3, 142–156, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.140.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H., Dejean, T., 2018. Environmental dna reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Sci. Rep. 8 (10361).

Prié, V., Danet, A., Valentini, A., Lopes-Lima, M., Taberlet, P., Besnard, A., Roset, N., Gargominy, O., Dejean, T., 2023. Conservation assessment based on large-scale monitoring of edna: Application to freshwater mussels. Biol. Cons. 283, 110089.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.

Rashid, Z.A., Amal, M.N.A., Shohaimi, S., 2018. Water quality influences on fish occurrences in sungai pahang, maran district, pahang, malaysia. Sains Malays. 47, 1941–1951.

Revenga, C., Campbell, I., Abell, R., de Villiers, P., Bryer, M., 2005. Prospects for monitoring freshwater ecosystems towards the 2010 targets. Phil. Trans. R. Soc. B 360, 397–413.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. Vsearch: a versatile open source tool for metagenomics. PeerJ 4, e2584.

Ruppert, K.M., Kline, R.J., Rahman, M.S., 2019. Past, present, and future perspectives of environmental dna (edna) metabarcoding: A systematic review in methods, monitoring, and applications of global edna. Glob. Ecol. Conserv. 17, e00547.

Salgado, J., Shurin, J.B., Vélez, M.I., Link, A., Lopera-Congote, L., González-Arango, C., Jaramillo, F., Åhlén, I., de Luna, G., 2022. Causes and consequences of recent degradation of the magdalena river basin, colombia. Limnol. Ocean. Lett. 7, 451–465.

Schiemer, F., 2000. Fish as indicators for the assessment of the ecological integrity of large rivers. Hydrobiologia 422–423, 271–278.

Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J., Rango, A., 2002. Remote sensing in hydrology. Adv. Water Resour. 25, 1367–1385, URL: https://www.sciencedirect.com/science/article/pii/S0309170802000659.

Schmutz, S., Cowx, I.G., Haidvogl, G., Pont, D., 2007. Fish-based methods for assessing european running waters: a synthesis. Fish. Manag. Ecol. 14, 369–380.

Schnell, I.B., Bohmann, K., Gilbert, M.T.P., 2015. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. Mol. Ecol. Resour. 15, 1289–1303.

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H.P., Wofsy, S.C., Xu, L., 2006. On the use of modis evi to assess gross primary productivity of north american ecosystems. J. Geophys. Res.: Biogeosciences 111.

Spear, M.J., Embke, H.S., Krysan, P.J., Vander Zanden, M.J., 2021. Application of edna as a tool for assessing fish population abundance. Environ. DNA 3, 83–91, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.94.

Spellerberg, I., 2008. Shannon–wiener index. In: Jørgensen, S.E., Fath, B.D. (Eds.), Encyclopedia of Ecology. Academic Press, Oxford, pp. 3249–3252, URL: https://www.sciencedirect.com/science/article/pii/B9780080454054001324.

Stevens, G.C., 1989. The latitudinal gradient in geographical range: How so many species coexist in the tropics. Amer. Nat. 133, 240–256, URL: https://doi.org/10.1086/284913.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., Brosse, S., 2021. Human impacts on global freshwater fish biodiversity. Science 371, 835–838, URL: https://www.science.org/doi/abs/10.1126/science.abd3369.

Sutela, T., Vehanen, T., Jounela, P., 2010. Response of fish assemblages to water quality in boreal rivers. Hydrobiologia 641, 1–10.

Tabacchi, E., Lambs, L., Guilloy, H., Planty-Tabacchi, A.-M., Muller, E., Décamps, H., 2000. Impacts of riparian vegetation on hydrological processes. Hydrol. Process. 14, 2959–2976.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. Trends Ecol. Evolut. 18, 306–314.

Val, P., Lyons, N.J., Gasparini, N., Willenbring, J.K., Albert, J.S., 2022. Landscape evolution as a diversification driver in freshwater fishes. Front. Ecol. Evol. 9.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Brun, M.L., Møller, P.R., Willerslev, E., Dejean, T., 2016. Next-generation monitoring of aquatic biodiversity using environmental dna metabarcoding. Mol. Ecol. 25, 929–942.

Vanhellemont, Q., 2020a. Automated water surface temperature retrieval from landsat 8/tirs. Remote Sens. Environ. 237, 111518, URL: https://www.sciencedirect.com/science/article/pii/S0034425719305371.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37, 130–137, URL: https://doi.org/10.1139/f80-017.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (1686).

Wolter, C., 2007. Temperature influence on the fish assemblage structure in a large lowland river, the lower oder river, germany. Ecol. Freshw. Fish 16, 493–503.

Yates, M.C., Fraser, D.J., Derry, A.M., 2019. Meta-analysis supports further refinement of edna for monitoring aquatic species-specific abundance in nature. Environ. DNA 1, 5–13, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.7.

Zong, S., Brantschen, J., Zhang, X., Albouy, C., Valentini, A., Zhang, H., Altermatt, F., Pellissier, L., 2023. Combining environmental dna with remote sensing variables to map fish species distributions along a large river. Remote. Sens. Ecol. Conserv..