Combining environmental DNA with remote sensing variables to map fish species distributions along a large river

Journal Article

Author(s):
Zong, Shuo; Brantschen, Jeanine; Zhang, Xiaowei; Albouy, Camille; Valentini, Alice; Zhang, Heng; Altermatt, Florian (iD); Pellissier, Loïc

RESEARCH ARTICLE

Combining environmental DNA with remote sensing variables to map fish species distributions along a large river

Shuo Zong[1,2] (iD), Jeanine Brantschen[3,4], Xiaowei Zhang[5] (iD), Camille Albouy[1,2] (iD), Alice Valentini[6], Heng Zhang[3,4] (iD), Florian Altermatt[3,4] (iD) & Loïc Pellissier[1,2] (iD)

[1]Department of Environmental Systems Science, Ecosystems and Landscape Evolution, Institute of Terrestrial Ecosystems, ETH Zürich, Zürich 8092, Switzerland
[2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf 8903, Switzerland
[3]Department of Aquatic Ecology, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, Dübendorf 8600, Switzerland
[4]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Winterthurerstrasse 190, 8057, Zürich, Switzerland
[5]State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, P. R. China
[6]SPYGEN, Le Bourget-du-Lac, France

Abstract

Biodiversity loss in river ecosystems is much faster and more severe than in terrestrial systems, and spatial conservation and restoration plans are needed to halt this erosion. Reliable and highly resolved data on the state of and change in biodiversity and species distributions are critical for effective measures. However, high-resolution maps of fish distribution remain limited for large riverine systems. Coupling data from global satellite sensors with broad-scale environmental DNA (eDNA) and machine learning could enable rapid and precise mapping of the distribution of river organisms. Here, we investigated the potential for combining these methods using a fish eDNA dataset from 110 sites sampled along the full length of the Rhone River in Switzerland and France. Using Sentinel 2 and Landsat 8 images, we generated a set of ecological variables describing both the aquatic and the terrestrial habitats surrounding the river corridor. We combined these variables with eDNA-based presence and absence data on 29 fish species and used three machine-learning models to assess environmental suitability for these species. Most models showed good performance, indicating that ecological variables derived from remote sensing can approximate the ecological determinants of fish species distributions, but water-derived variables had stronger associations than the terrestrial variables surrounding the river. The species range mapping indicated a significant transition in the species occupancy along the Rhone, from its source in the Swiss Alps to outlet into the Mediterranean Sea in southern France. Our study demonstrates the feasibility of combining remote sensing and eDNA to map species distributions in a large river. This method can be expanded to any large river to support conservation schemes.

Introduction

To counter global biodiversity erosion, ecosystem governance and management require an increase in the speed, accuracy and ease of biodiversity data collection and processing (Dornelas et al., 2019; Makiola et al., 2020). This entails a shift from expert monitoring towards high-throughput data acquisition technology (Cordier

et al., 2019). Freshwater ecosystems have been experiencing major declines in biodiversity (Darwall et al., 2018; Grooten et al., 2018) as a result of river canalization, dam building, pollution and over-exploitation (Maavara et al., 2020). To reverse these trends and effectively manage freshwater ecosystems, reliable data-driven conservation and policy-oriented planning are needed (Darwall et al., 2018). Fortunately, our ability to rapidly monitor the status of freshwater ecosystems to inform policies has increased with recent technological developments. On the one hand, fast inventories of entire freshwater communities are now possible with the emergence of environmental genomics and specifically environmental DNA (eDNA) (Bohmann et al., 2014; Cordier, 2020; Deiner et al., 2017; Pawlowski et al., 2018; Thomsen & Willerslev, 2015). On the other hand, the increased spatial and temporal resolution of satellite observations allows regular surveys of ecosystems, which can be used with machine learning to reconstruct explicit maps of species distribution when combined with point-based species information (Barbarossa et al., 2020; Vihervaara et al., 2017). The resulting species distribution maps can inform conservation planning and spatial prioritization (Ferrier, 2002; Jetz et al., 2012). The combination of eDNA and remote sensing could enable the fast production of fish species distribution maps, especially in regions where little species information is currently available (Pettorelli et al., 2014; Wang & Gamon, 2019; Yamasaki et al., 2017). However, the feasibility and performance of such data and method associations require evaluation.

Environmental DNA is becoming a method of choice to survey freshwater species at large spatial scales (Altermatt et al., 2020; Deiner et al., 2017; Lyet et al., 2021), as it can be applied to large river corridors (Blackman et al., 2021; Carraro et al., 2020). eDNA is genetic material obtained directly from environmental samples (e.g. soil, sediment and water), and it is characterized by a complex mixture of intracellular (from living cells) or extracellular DNA (originating from skin, hair, urine, faeces or carcasses) (Pawlowski et al., 2020; Taberlet et al., 2012). eDNA thus offers an integrative view of ecosystem composition (Deiner et al., 2017; Ficetola et al., 2008). When coupled with high-throughput DNA sequencing methods, eDNA 'metabarcoding' can contribute to the rapid assessment and monitoring of species distributions across all levels of life, from prokaryotes to eukaryotes, with a higher detection capacity and cost-effectiveness than traditional methods (e.g., Polanco Fernández et al., 2021). The sequences from the high-throughput sequencing are then compared with reference barcode libraries to establish taxonomic lists directly from environmental samples (Taberlet et al., 2012). Ultimately, these lists can be used to assess species occupancy, but

also ecosystem functioning and health status (Cordier et al., 2019). An increasing number of initiatives propose to use eDNA metabarcoding routinely and globally to monitor ecosystems (Berry et al., 2021). Large rivers, defined here as rivers at least 60 m wide, deliver many ecosystem services to human populations living on their banks (Palmer et al., 2008), but are under increasing anthropogenic pressure, for example relating to damming, pollution and overfishing (Maavara et al., 2020). In this context, eDNA offers a cost-efficient method for monitoring species within these large rivers (Pont et al., 2018). Recently, broad-scale applications of eDNA along the Maroni and Oyapock Rivers in South America (Cantera et al., 2020), the Yangtze River in China (Li et al., 2018; Zhang et al., 2022) and the Rhone River in Switzerland and France (Pont et al., 2018) have demonstrated the reliability of eDNA in detecting fish occupancy at multiple locations along the river corridor. However, species monitoring with eDNA remains spatially limited, often with tens of kilometres between two sampling sites (Blackman et al., 2021). These patchy data could be associated with remotely sensed information about the habitat to build continuous spatial models of species distributions.

Remote sensing involves regular measurements over time (Ozesmi & Bauer, 2002; Reul et al., 2020), matching the spatial and temporal resolution of eDNA metabarcoding data. Products from remote sensing have led to a more precise characterization of the ecological properties of freshwater ecosystems (Campbell et al., 2011; Castello & Macedo, 2016; Revenga et al., 2005), including the water bodies themselves (Watanabe et al., 2015) and the surrounding terrestrial environments (Growns et al., 2003; Karra et al., 2021). Rivers generally occupy a small spatial fraction of the landscape, and thus, their characterization requires the use of high-resolution images (Kuhn et al., 2019). High-resolution remote sensing data provide accurate information on the narrow features of river systems and could therefore help explain the distribution of species detected in eDNA signals. In river systems, temperature is expected to be a major factor determining species distributions for many taxonomic groups. In particular, it determines the spatial distribution of many fish species (Brazner et al., 2005; Buisson et al., 2008; Heino et al., 2002; Shuter & Post, 1990). Since fishes are ectothermic animals, they are dependent on their thermal environment for survival, breeding and developmental success (Gillooly et al., 2001; Mann & Blackburn, 1991; Mills & Mann, 1985; Nunn et al., 2003; Wolter, 2007). In addition, other environmental factors influence their distribution and community composition, including water quality (Watanabe et al., 2015), water clarity (Harrington et al., 1992; Lee et al., 2016), substrate type (Humpl & Pivnička, 2006) and the structure of the

riparian vegetation (Growns et al., 2003; Maridet et al., 1998), which can all be approximated by remote sensing information (Bae et al., 2019; Bergen et al., 2009; Drusch et al., 2012; Turner et al., 2003).

Species distribution models (hereafter: SDMs) relate species observations to descriptors of the environment and produce response curves indicating species responses to environmental gradients. By matching eDNA and remote sensing data, SDMs can provide insight into species responses to environmental gradients and also into species distributions if the spatial layers truly reflect features related to the ecological preferences of species (Randin et al., 2020). By using SDMs that integrate a set of machine learning methods, species occurrence data (e.g. from eDNA metabarcoding detection) and spatial variables (e.g. remote sensing data), it is possible to predict species' potential occupancy over a landscape (Guisan & Zimmermann, 2000). In freshwater ecology and conservation, SDM maps have been used to design protected areas (Esselman & Allan, 2011) and to predict the risk of species extinction under climate change (Bond et al., 2011), but these applications have been limited to rivers with existing species surveys. Hence, the rapid generation of species distribution maps through the combination of eDNA, remote sensing and machine learning would be especially valuable in the management of regions where other species information is lacking. However, the source of the transported eDNA is uncertain (from local to several kilometres away) (Carraro et al., 2020), and it is essential to consider this uncertainty when associating the species occupancy data from eDNA with the remote sensing variables to integrate their ecological signals (Goldberg et al., 2015). Moreover, the scarcity of continuous freshwater-specific environmental predictors has been a major challenge in efforts to produce robust estimations of the geographic habitat suitability of freshwater species by means of SDMs (Domisch et al., 2015). Hence, cloud processing of high-resolution remote sensing predictors for riverine systems could represent a major tool for the development of freshwater SDMs (Amani et al., 2020; Gorelick et al., 2017). Combining eDNA with the remote sensing cloud infrastructure could offer a high-throughput support tool for conservation decisions about large river systems.

Here, we investigated whether eDNA metabarcoding data can be coupled with remote sensing predictors, matched both spatially and temporally, to map fish species distributions in large river systems. To generate ecological variables of freshwater habitats and the surrounding vegetation, we used a combination of imagery from Landsat 8, Sentinel 2 and associated derived products. The resulting set of variables represented water temperature, water quality, vegetation type surrounding rivers (Growns et al., 2003), terrestrial properties and human activities

based on the Google Earth Engine cloud computing platform (Gorelick et al., 2017; Kennedy et al., 2019). We focused on the Rhone River, where an intense eDNA sampling campaign was completed to inventory the fish composition of 198 samples in Switzerland and France. Given that the signal from eDNA is partly carried downstream by the current, relating sampling locations to ecological variables requires the selection of an appropriate spatial scale for the eDNA signal. To consider this issue, we evaluated the influence of applying different spatial scales (5, 10, 15 and 20 km) for the environmental extractions. Specifically, we asked the following questions:

1. Can the species occupancy data derived from eDNA detection be combined with environmental variables from remote sensing to capture species' ecological preferences and map their potential distribution along a large river corridor?
2. What are the most important environmental variables, derived from remote sensing, for predicting fish species distributions along the Rhone River?
3. How do the spatial scale and direction of the extraction window of remote sensing layers surrounding the eDNA sampling point (5, 10, 15 and 20 km in the upstream direction) influence the performance of SDMs derived from eDNA, and what is the optimal scale to use?

Materials and Methods

eDNA datasets

Along the whole length of the Rhone River, starting in its alpine upstream area in Switzerland and ending at its outlet into the Mediterranean Sea in France, eDNA samples were taken at a total of 110 sites (Pont et al., 2018). Given the river configuration, more samples were collected in France than in Switzerland, and the confluences were larger in France. While the sampling, extraction and sequencing protocols also differed between the samples from Switzerland and France, the bioinformatic analyses resulted in qualitatively comparable data at the presence/absence level when restricted to the most common species (field and laboratory work details in Supporting Information). All lake species were excluded, and only species present in both databases were included in the final curated dataset.

Generation of remote sensing variables

From the MultiSpectral Instrument (MSI) onboard Sentinel 2, we used the Level 1C (Top-of-Atmosphere Reflectance) and Level 2A (Surface Reflectance) products (Drusch et al., 2012). We computed variables representing

water-quality characteristics. Specifically, we computed Secchi depth (SD), trophic state index (TSI) and chlorophyll-*a* concentration (Chl*a*) using the approaches developed by Page et al. (2018) (Gessesse & Melesse, 2019), which are the most commonly used metrics for water-quality measurements (Olmanson et al., 2008).

We derived river water surface temperature (RST) from the thermal infrared sensor onboard Landsat 8 (details in Supporting Information). We employed several vegetation indices to reflect the multiple properties of the vegetation surrounding the river. We computed the normalized difference vegetation index (NDVI), from the Landsat 8 Operational Land Imager, to represent the vegetation conditions surrounding the river. Low NDVI values indicate moisture-stressed vegetation, while higher values indicate a higher density of green vegetation (Gessesse & Melesse, 2019). We further considered the enhanced vegetation index (Sims et al., 2006), leaf area index (Zheng & Moskal, 2009) and gross primary productivity (GPP) from MODIS products to represent the structure of the vegetation (Wulder et al., 1998) and its productivity. Moreover, we considered forest canopy height, calculated following the methods of Lang et al. (2022), to investigate its impact on the habitat of fish species.

The morphology of a river can influence the hydrology and other ecological conditions, and in turn the freshwater species distribution. For example, the river channel width and shape can influence the occurrence of important habitats for fish (Kail et al., 2015). We therefore considered the surrounding water area and the river width to represent the morphology of each river pixel. We computed the surrounding water area by summing all the water pixels within pre-defined surrounding buffers (see "Extraction windows of remote sensing variables" section). We acquired river width data from the Global River Width from Landsat data produced by Allen and Pavelsky (2018). We further used terrain factors representing the channel morphology and the form of the river banks, which are expected to influence the species distribution pattern in rivers. We computed the slope of the terrain surrounding the river and the height difference between the bank and the water surface, which can serve as proxies for species habitats. Finally, as human activities surrounding the river can influence freshwater species distributions, we acquired the human modification index (Kennedy et al., 2019) and computed its median value within the pre-defined buffers.

Extraction windows of remote sensing variables

For each eDNA sample location, we extracted the values of the remote sensing variables using a neighbourhood analysis method defined by a spatial buffer around the sampling point (Fig. 1), considering both standard and upstream-directed buffers. We separated the remote sensing variables into river (blue) and riparian (green) variables. For blue variables (e.g. RST and water clarity), we masked the non-water pixels, whereas for green variables (canopy height and NDVI), we masked the water pixels. We first considered circular buffers of 5, 10, 15 or 20 km radius centred around each eDNA sampling site. Second, to accommodate the direction of the river flow and the potential dispersion of eDNA molecules, we considered upstream-directed circular buffers from the eDNA sampling sites. For this, we designed the circular buffers (5, 10, 15 or 20 km radius) to be centred 5, 10, 15 or 20 km upstream such that the edge of the buffer passed through the eDNA sampling site. We extracted the remote sensing values and computed medians and standard deviations of the values extracted within the buffers. In total, we generated 23 variables, extracted at four different scales with two buffer types.

Species distribution modelling

We generated habitat suitability maps using SDMs (Guisan & Zimmermann, 2000) for 29 fish species detected in 198 eDNA samples from the 110 sampling sites. We only considered species with at least 20 occurrences for building the models. We first investigated the correlations between the remote sensing variables. We excluded variables that were highly correlated with each other (Pearson's correlation >0.7) (Dormann et al., 2013; Hirzel et al., 2006; Petitpierre et al., 2017; Zurell et al., 2020), and we ranked the variables based on their correlation with each target species individually and then extended these rankings to create a global ranking across all species. We kept those variables with higher univariate correlation global rank in the highly correlated pairs. We retained 14 variables for further modelling (Table 1).

We related species presence and absence to environmental conditions using generalized linear models (GLMs; McCullagh & Nelder, 2019), gradient boosting machines (GBMs; Friedman, 2002; Natekin & Knoll, 2013) and random forests (RFs; Liaw & Wiener, 2002). For each model, we applied a simple parametrization, as defined by Brun et al. (2020). We used recursive feature elimination (Granitto et al., 2006) to first reduce the number of variables so that there were at least 10 occurrences for each variable considered in the model (Breiner et al., 2018). Next, we fitted different models for the different fish species, depending on their number of occurrences along the whole river. We ran the standard models for species with balanced numbers of presences and absences. By contrast, for species with fewer than 60 or more than 140 presence
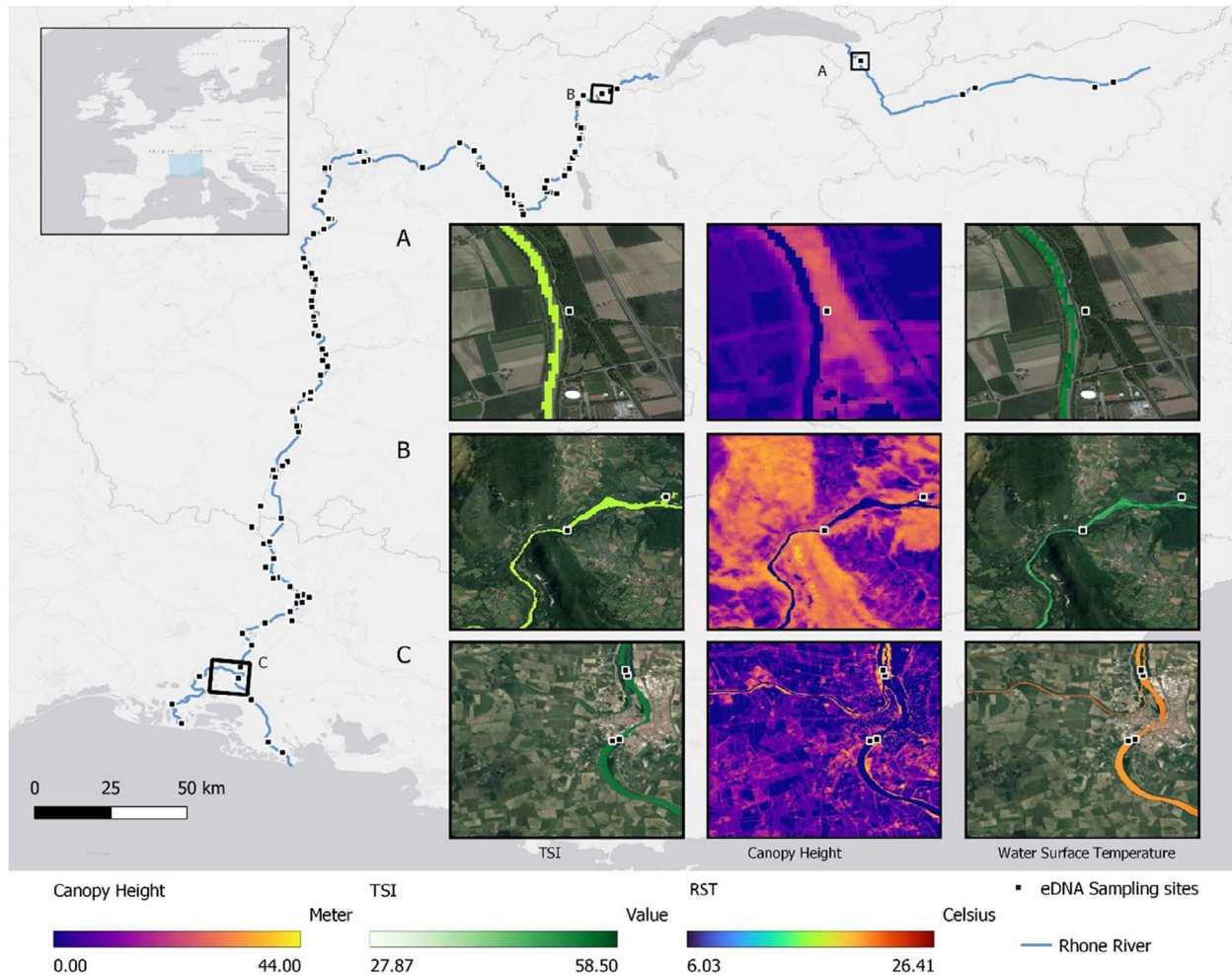
Figure 1. Map of environmental DNA (eDNA) sampling sites along the full extent of the Rhone River in Switzerland and France (background) and the values of three remote sensing variables [inset maps; trophic state index (TSI), canopy height and river water surface temperature (RST)] at three sites (A, B, C). To aid visualization, A, B and C have different scales, and the three variables have different colourmaps.

observations, we used the method of ensemble of small models (Lomba et al., 2010).

We conducted a fivefold cross-validation for all the models (including both small models and regular models) to calculate the true skill statistic (TSS; Allouche et al., 2006), after which we retained only models with a TSS >0.4 and with better performance than null models. To compute null models, we randomly resampled the presence and absence 100 times from 110 sites using bootstrapping for each species (details are given in the Supporting Information) (Raes & ter Steege, 2007; van Proosdij et al., 2016). We compared the upper 95% confidence interval of the null models' performance with the average performance of SDMs (Raes & ter Steege, 2007). We computed the median effective TSS (median TSS difference between SDMs and null models) to indicate the

model performance. We compared the effective TSS across the buffer scales and types used for the extraction of the remote sensing variables.

We produced an ensemble model for each species, each statistical method, and each buffer scale and type, taking the median of the predicted suitability of the single models (details of modelling and model performance calculations following a standard protocol are given in the Supporting Information). We converted model-based projections to binary presence/absence using the threshold of maximum TSS (Allouche et al., 2006). We considered the species to be present in areas where at least two of the three ensembles predicted presence. In total, we ran 696 SDMs, representing 29 species, using three machine learning methods from eight combinations of buffer scales and types. We compared the predictive power of the

Table 1. Fourteen remote sensing variables considered in the species distribution models.

Abbreviation	Variable name	Description
RST	Median river water surface temperature (Vanhellemont, 2020)	Median value of river water surface temperature within the buffer
TSI	Median trophic state index (Page et al., 2018)	Median value of trophic state index within the buffer
SD-Med	Median Secchi depth (Page et al., 2018)	Median value of Secchi depth within the buffer
SD-Std	Standard deviation Secchi depth (Page et al., 2018)	Standard deviation of Secchi depth within the buffer
Chla	Median chlorophyll-*a* (Cannizzaro & Carder, 2006; Page et al., 2018)	Median value of chlorophyll-*a* concentration within the buffer
NDVI	Median NDVI (Gessesse & Melesse, 2019)	Median value of normalized difference vegetation index within the buffer
CanopyH	Standard deviation canopy height (Lang et al., 2022)	Standard deviation of canopy height within the buffer
EVI	Standard deviation EVI (Sims et al., 2006)	Standard deviation of enhanced vegetation index within the buffer
GPP	Standard deviation GPP (Wulder et al., 1998)	Standard deviation of gross primary productivity within the buffer
Slope-Med	Median slope (Farr et al., 2007)	Median value of surrounding terrain slope within the buffer
Slope-Std	Standard deviation slope (Farr et al., 2007)	Standard deviation of surrounding terrain slope within the buffer
HumModi	Median human modification index (Kennedy et al., 2019)	Median value of human modification index within the buffer
RW	Median river width (Allen & Pavelsky, 2018; Kail et al., 2015)	Median value of river width within
WaterA	Sum of surrounding water area	Total water area within the buffer

different model types (GLMS, GBMs and RFs) and of models with different buffer types and sizes and with different fish species, testing the significance with a linear model associated with an analysis of variance.

Importance of remote sensing variables in SDMs

To assess the importance of the remote sensing variables in determining the species distributions, we recorded the number of times a predictor was retained in the well-performing models (with TSS >0.4 and better performance than null models) for each species. In addition, we computed the importance of the predictor using a randomization procedure. For each of the well-performing models, we further employed a randomized approach to assess the importance of each variable in predicting the outcome, using the R package BIOMOD2 (Thuiller et al., 2023). The method involved randomly shuffling values to each variable and systematically repeating the process for all variables. Through this process, we quantified the loss of predictive power as measured by the correlation between the original results and the shuffled ones.

Results

Species recovered from eDNA metabarcoding

We recovered 71 fish species from 51 genera from the eDNA metabarcoding data. After removing species associated with lakes and those that are rare or too widespread, we retained 29 species for species distribution modelling. Specifically, of all the species detected from the 198 eDNA samples, two species had >180 detections and were thus considered too widespread for modelling their associations with environmental variables. Furthermore, 20 species had ≤20 occurrences, which is not sufficient to model their distribution (Breiner et al., 2015). The full list of detected species and their number of occurrences can be found in Table S1.

Spatial patterns of remote sensing variables

The longitudinal profile of the environmental variables derived from remote sensing showed contrasting spatial patterns along the Rhone River (Fig. 2). RST was generally low in the Rhone valley of Switzerland and gradually decreased further before Lake Geneva, followed by a steep increase from Lake Geneva (128–210 km of the river) to 410 km. After that point, a trough in RST occurred and the values became more even until the Mediterranean Sea. The longitudinal TSI profile showed high values at the source in Switzerland, reflecting the relatively high nutrient concentrations there, and a decline after kilometre 200. The profile fluctuated, with several peaks and troughs occurring from Lake Geneva to the Mediterranean Sea. Similarly, SD decreased after Lake Geneva, indicating that transparency decreased towards the river mouth. The terrain slope to the river had two peaks, at about kilometres 250 and 450. The longitudinal profile of GPP showed initially high values reaching a peak near kilometre 290, followed by a decrease towards the mouth
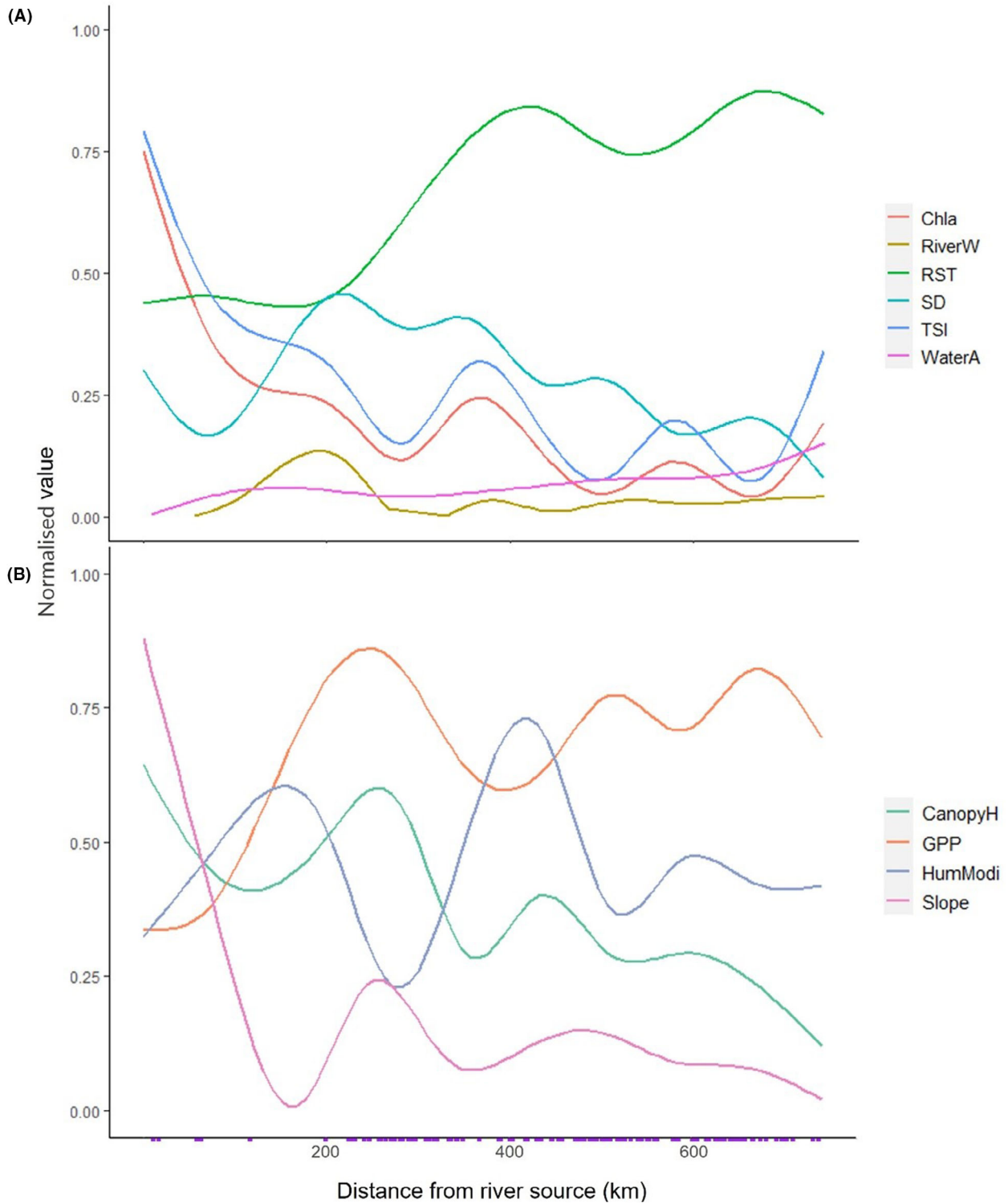
Figure 2. Longitudinal profiles of normalized remote sensing variables. (A) Longitudinal profiles of the blue (river) variables chlorophyll-*a* concentration (Chla), river width (RiverW), river water surface temperature (RST), Secchi depth (SD), trophic state index (TSI) and surrounding water area (WaterA). (B) Longitudinal profiles of the green (riparian) variables forest canopy height (CanopyH), gross primary productivity (GPP), human modification index (HumModi) and terrain slope. Note that the values shown here are the medians of the values within the buffers. The purple ticks on the x-axis indicate the sample locations along the Rhone river.

of the Mediterranean Sea. Overall, the longitudinal profiles emphasized the marked gradients in the variables derived from remote sensing along the course of the Rhone.

Species distribution modelling

The SDMs built from fish species occurrences detected in eDNA and environmental variables derived from remote sensing performed well in general and captured some of the species distribution patterns successfully (Figures S2 and S4–S8). The overall TSS of all 29 species was >0.4 (mean 0.58, std. 0.22), which is considered the minimum for informative models (Allouche et al., 2006). Moreover, the SDMs performed significantly better than the upper 95% confidence interval of 100 times randomly generated null models (28 out of 29 species significantly better), which indicates that the models performed better than null models (Figures S6–S8). When comparing the three model types, we found that the TSS values of RF models were generally higher than values of the two other modelling methods considered (Figure S2). Some species, including *Gasterosteus aculeatus* and *Esox lucius*, preferred moderately productive, clear and cold water and high vegetation coverage, as found in the upper Rhone. By contrast, a few species were mainly distributed in warmer

and more turbid waters in the southern part of the Rhone, such as *Silurus glanis* (Fig. 3) and *Blicca bjoerkna*. Other species were found in waters of intermediate RST (>17 °C), distributed along the French Rhone from Lake Geneva to the Mediterranean Sea, such as *Cyprinus carpio* (Fig. 3) and *Rhodeus amarus*.

## Effect of the scale of remote sensing variables

Models with an upstream-directed buffer had a significantly higher predictive power than models with the sampling point at the centre of the buffer (median effective TSS difference = 0.009, $P < 0.01$; Figure S3). Model performance additionally decreased as buffer size increased, with a significant difference between the 5 and 20-km buffers (median effective TSS difference = 0.026, $P < 0.001$; Figure S3).

Importance of variables

The variables derived from remote sensing had different explanatory powers across the modelled species in the GLMs, GBMs and RFs (Fig. 4). The five most important variables were RST (blue), SD (blue), terrain slope (slope, green), TSI (blue) and GPP (green). The comparison of
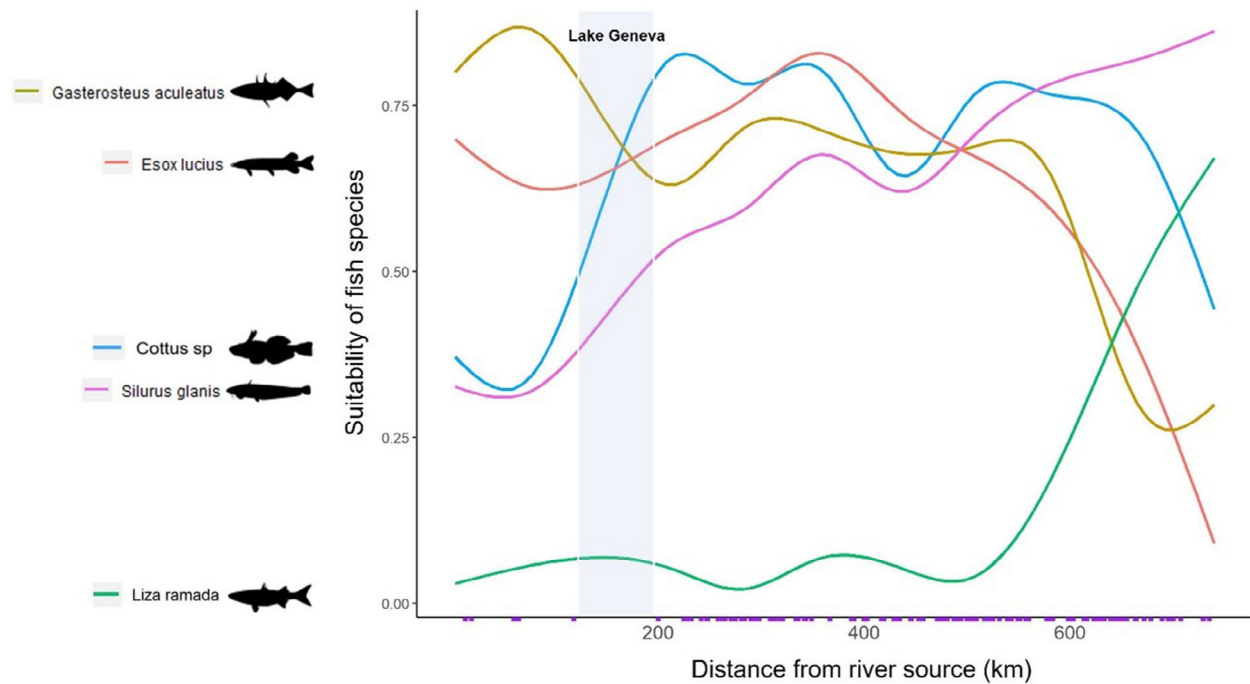
Figure 3. Longitudinal profiles of selected species (Cottus sp., *Esox lucius*, *Gasterosteus aculeatus*, *Liza ramada* and *Silurus glanis*). The suitabilities of the fish species are the predictions from the species distribution models. The purple ticks on the x-axis indicate the sample locations along the Rhone River.
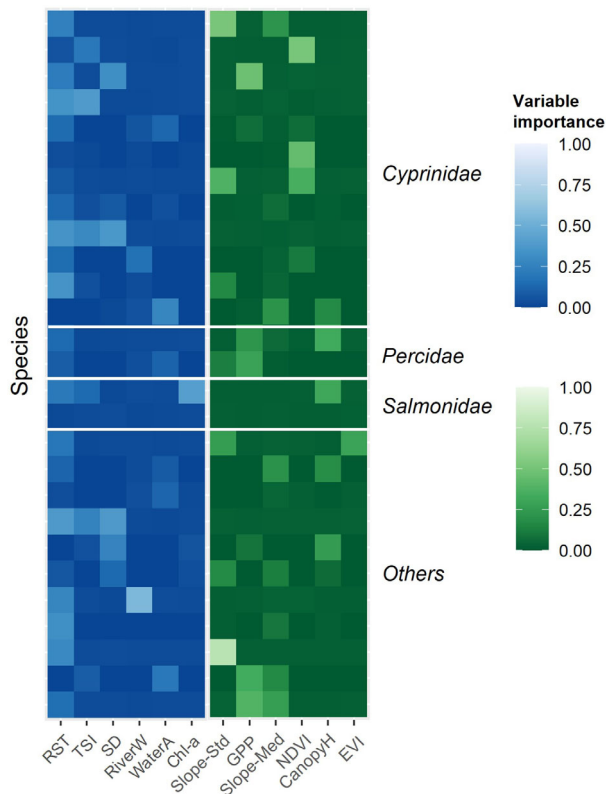
Figure 4. Variable importance in all the models (top 12 most important variables). The heatmap shows the variable importance across all the qualified models for all the species. The pixel values indicate the variable importance within each species distribution model (SDM). The rows correspond to species, labelled by their family names respectively (Cyprinidae, Percidae and Salmonidae have more than one species, and are labelled with family names, the other species are labelled as Others). The columns represent the variables, with blue indicating river variables and green indicating riparian variables. The river and riparian variables, from left to the right in the columns, represents decreasing level of importance. CanopyH, standard deviation of canopy height within the buffer; EVI, standard deviation of enhanced vegetation index within the buffer; GPP, standard deviation of gross primary productivity within the buffer; NDVI, median of normalized difference vegetation index within the buffer; RST, median of river water surface temperature within the buffer; SD, median of river water Secchi depth within the buffer; Slope-Med, median of terrain slope within the buffer; Slope-Std, standard deviation of terrain slope within the buffer; TSI, median of trophic state index of river water body within the buffer; WaterA, total water area within the buffer.

the different machine-learning models indicated good agreement in the ranking of variable importance. RST was the most important variable across most models, irrespective of the modelling method. We found some variation among the fish species regarding the most important variables in the models. While RST, slope and SD were the most important variables explaining the distribution

of most species, a few species distributions were mainly associated with green variables, including GPP and NDVI. Indeed, both blue and green variables drove the distributions along the Rhone River in our models of the different species.

Discussion

Our study demonstrates that eDNA metabarcoding data can be coupled with environmental variables derived from remote sensing to adequately predict species distributions along a long river corridor. Our findings illustrate the potential of satellite sensors to scale up eDNA data in aquatic ecosystems, as previously achieved in soil eDNA studies (Lin et al., 2021). Our results show that eDNA can be seamlessly combined with remote sensing data to calibrate response curves and map the distribution of fish species along an entire river corridor. Precise maps derived from remote sensing data can therefore help to refine conservation strategies and support the monitoring of the environmental conditions critical for maintaining biodiversity along large rivers (Asner et al., 2022; Esselman & Allan, 2011). Our study fills major gaps in essential biodiversity variables by providing remotely sensed indicators for the distribution of critical fish species (Asner et al., 2022; Turak et al., 2017).

The new generation of satellite sensors enables the computation of variables that are generally more ecologically meaningful for mapping species distributions (Nagendra, 2001; Randin et al., 2020). The use of remote sensing for mapping terrestrial organisms has been developed in the last 20 years (He et al., 2015), and data derived from remote sensing have been shown to contain relevant ecological information for mapping species distributions (Schwager & Berg, 2021). Remote sensing has recently been applied to model aquatic vegetation (Rowan & Kalacska, 2021) and to map the typology of coral reefs (Asner et al., 2022). Here, we demonstrate the potential of using remote sensing variables describing river environments in predictions of fish species distributions. The application of remote sensing to rivers is non-trivial because they occupy a small fraction of the landscape and a high spatial resolution is necessary, for instance the 10–20 m resolution of Sentinel 2. Our results further suggest that direct water-related variables are more critical than riparian ones, indicating the power and importance of mapping water properties for river ecosystems globally (Markovic et al., 2012). So far, most river applications of remote sensing have been focused on the riparian vegetation in terms of predictors (Morgan et al., 2020), while the modelling of aquatic organisms that are critically endangered requires further development.

Water temperature and water quality are generally the most important drivers of species distributions in freshwater ecosystems (Bouska et al., 2015; Knudby et al., 2010; Olden & Jackson, 2002; Sandström et al., 2016; Vezza et al., 2015). In our analyses combining eDNA and remote sensing variables, RST and water clarity were the most important variables driving the species distributions. Surface water temperature is a well-known determinant of fish species distributions in relation to fish physiological properties (Farkas et al., 2001). The water clarity indicator may reflect the level of productivity in the water, an indication of food resource abundance, but may also correspond to the level of water pollution (Hoyer et al., 2002). Our validation of RST with river monitoring station data demonstrated a very high level of agreement (Figure S12). However, in future eDNA campaigns, with water quality and clarity properties also collected, the other water-related variables can also be validated, and this will support future aquatic eDNA studies and further improve the accuracy of our analysis. Green variables (GPP, NDVI, slope and canopy height) should complement hydrological variables by documenting the riparian habitat of fish species, but they generally showed lower importance in our models, except for a few species. Pike (E. lucius) preferred moderate productivity in clear and cold water and riparian areas with high vegetation coverage, as found in the upper Rhone, which is in line with previous knowledge of this species. The reduction in macrophytes caused by eutrophication has a harmful impact on pike populations as these plants are essential for the successful reproduction and survival of young pike (Bry, 1996). Moreover, the low visibility and lack of cover resulting from eutrophication create unfavourable hunting conditions for pike (Bry, 1996; Zarkami, 2008). The model prediction of pike presence in cold water also corresponds to habitats with high oxygen concentrations, which favour the survival of larvae and embryos (Siefert et al., 1973). This example highlights how remote sensing variables can capture the ecological conditions required for fish species, facilitating the modelling and mapping of their distributions.

Our modelling approach identified contrasting distributions of fish species along the Rhone River and in relation to distinct environmental conditions. For instance, we found that the European catfish (S. glanis) is distributed in warm and turbid waters in the southern part of the Rhone. This result is in accordance with previous studies that have identified the low oxygen requirement of this species (Lelek, 1987), as it has an affinity for high temperatures (physiological optimum about 25–27°C) (Copp et al., 2009), its relative tolerance of pollution (Lelek, 1987), and its presence in large rivers and in coastal areas with low salinity (<15) (Copp et al., 2009).

We additionally detected major shifts in species distributions along the Rhone River (Pont et al., 2015). Specifically, Lake Geneva is associated with strong changes in water conditions, reshaping the properties of the river and thus the species composition. While some species, such as Gasterosteus aculeatus, are mainly found in the colder stretch near the source of the river, other species, including S. glanis, are found more often in the warmer waters downstream. Overall, our study shows how the distribution of fish species can be recovered by combining eDNA metabarcoding and remote sensing variables.

The present study has several limitations related to the sampling and the uncertainty regarding eDNA transport along the river. The sampling in Switzerland was less extensive than in France, and while our modelling results were not sensitive to this difference (Figure S10), a random stratified sampling design would be more appropriate for species distribution modelling (Hirzel & Guisan, 2002). For instance, there are species known to be more common in the Swiss section of the Rhone, but due to the limited number of sampling sites, occurrences were too few to model their distribution. Thus, future eDNA sampling campaigns should be designed with the aim of spatial modelling and should include a reasonable number of sample sites across each of the environmental strata of the river. Moreover, in rivers, the eDNA collected at one point of the river does not originate from that exact location, but from an undefined area upstream. This might complicate the association between the species occupancy signal and remote sensing variables, and should be considered in modelling efforts (Carraro et al., 2021; Turner et al., 2014). Our results indicate that the best association between the eDNA signal and remote sensing data is at a 5-km resolution and in an upstream direction, as expected from the eDNA origin. These findings support those of Pont et al. (2018), where some effect of transport was detected but a complete turnover of the eDNA signal was found over distances of 10–130 km in the Rhone. Using a similar approach in a subalpine freshwater system with faster-running water, Zhang et al. (2023) found a spatial association between riverine eDNA diversity and remote sensing spectral diversity of terrestrial ecosystems upstream, peaking at a 400 m distance yet still detectable within a radius of up to 3.3 km. These differences between riverine systems indicate that the optimal shape and size of the buffer used to relate remote sensing information to eDNA might vary across rivers, depending on channel width and flow, and that sensitivity analyses should be performed before running models in new systems. Even in the same river system, the optimal shape and size of the buffer may vary among sections. Therefore, in future studies it may be

necessary to model eDNA transportation across each river section using variables such as river flow rate.

In conclusion, in this study we demonstrated the feasibility of coupling eDNA with remote sensing data to map the fish species distributions of large rivers. As freshwater species have not been inventoried for many large rivers, even as pressure from humans increases, our approach represents a way forward for rapidly mapping the species distribution status in such riverine ecosystems. Recent examples in the Maroni (Cantera et al., 2020) and the Yangtze (Zhang et al., 2019) have demonstrated that eDNA metabarcoding can be applied in the span of a few weeks to inventory the species distribution of entire river corridors. In turn, we show that the combination of rapid eDNA surveys with remote sensing data is feasible and informative regarding the spatial distribution of species. Beyond mapping single species, machine learning approaches could also be employed to map other biodiversity properties, such as species richness and phylogenetic or functional diversity. Moreover, with more frequent eDNA data collection in the future, our model could be applied across time and validated with temporal eDNA data to become more robust. Changes in the habitats of migratory fishes could even be predicted and validated. With the rise of cloud computation and more advanced machine learning techniques, long-term eDNA monitoring data coupled with remote sensing data could offer a powerful tool for near-direct mapping of changes in diversity over time. While our approach was applied to a large river, similar indicators for other aquatic systems, such as coastal reefs, are being developed. Hence, we foresee that many ecosystems in need of conservation will benefit from a targeted combination of remote sensing and eDNA monitoring.

Acknowledgements

Author Contributions

LP conceived the ideas and supervised SZ. JB provided and analyzed the Swiss Rhone section data. SZ computed the remote sensing variables, compiled the occurrence data from eDNA data, performed the models, and ran the analysis. SZ, CA and LP wrote the first version of the manuscript. All the authors SZ, JB, XZ, CA, AV, HZ, FA and LP contributed to the writing and the improvement of the manuscript.

Conflict of Interest

The authors declare there are no conflicts of interests.

References

Allen, G.H. & Pavelsky, T.M. (2018) Global extent of rivers and streams. *Science*, **361**(6402), 585–588.

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**(6), 1223–1232.

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**(5), 607–618.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M. et al. (2020) Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 5326–5350.

Asner, G.P., Vaughn, N.R., Martin, R.E., Foo, S.A., Heckler, J., Neilson, B.J. et al. (2022) Mapped coral mortality and refugia in an archipelago-scale marine heat wave. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(19), e2123331119.

Bae, S., Levick, S.R., Heidrich, L., Magdon, P., Leutner, B.F., Wöllauer, S. et al. (2019) Radar vision in the mapping of forest biodiversity from space. *Nature Communications*, **10**(1), 1–10.

Barbarossa, V., Schmitt, R.J., Huijbregts, M.A., Zarfl, C., King, H. & Schipper, A.M. (2020) Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(7), 3648–3655.

Bergen, K., Goetz, S., Dubayah, R., Henebry, G., Hunsaker, C., Imhoff, M. et al. (2009) Remote sensing of vegetation 3-D structure for biodiversity and habitat: review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*, **114**(G2), 883.

Berry, O., Jarman, S., Bissett, A., Hope, M., Paeper, C., Bessey, C. et al. (2021) Making environmental DNA (eDNA) biodiversity records globally accessible. *Environmental DNA*, **3**(4), 699–705.

Blackman, R.C., Osathanunkul, M., Brantschen, J., Di Muri, C., Harper, L.R., Mächler, E. et al. (2021) Mapping biodiversity hotspots of fish communities in subtropical streams through environmental DNA. *Scientific Reports*, **11**(1), 1–11.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M. et al. (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, **29**(6), 358–367.

Bond, N., Thomson, J., Reich, P. & Stein, J. (2011) Using species distribution models to infer potential climate change- induced range shifts of freshwater fish in south-eastern Australia. *Marine and Freshwater Research*, **62**(9), 1043–1061.

Bouska, K.L., Whitledge, G.W. & Lant, C. (2015) Development and evaluation of species distribution models for fourteen native central us fish species. *Hydrobiologia*, **747**(1), 159–176.

Brazner, J.C., Tanner, D.K., Detenbeck, N.E., Batterman, S.L., Stark, S.L., Jagger, L.A. et al. (2005) Regional, watershed, and site-specific environmental influences on fish assemblage structure and function in western Lake Superior tributaries. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**(6), 1254–1270.

Breiner, F.T., Guisan, A., Bergamini, A. & Nobis, M.P. (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**(10), 1210–1218.

Breiner, F.T., Nobis, M.P., Bergamini, A. & Guisan, A. (2018) Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, **9**(4), 802–808.

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R.O., Wang, Z. et al. (2020) Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, **47**(1), 130–142.

Bry, C. (1996) Role of vegetation in the life cycle of pike. In: Craig, J.F. (Ed.) *Pike*. Dordrecht: Springer, pp. 45–67.

Buisson, L., Thuiller, W., Lek, S., Lim, P. & Grenouillet, G. (2008) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*, **14**(10), 2232–2248.

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 Sensing of Environment*, **115**(9), 2402–2414.

Cannizzaro, J.P. & Carder, K.L. (2006) Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters. *Remote Sensing of Environment*, **101**(1), 13–24.

Cantera, I., Decotte, J.-B., Dejean, T., Murienne, J., Vigouroux, R., Valentini, A. et al. (2020) Characterizing the spatial signal of environmental DNA in river systems using a community ecology approach. *Molecular Ecology Resources*, **22**(4), 1274–1283.

Carraro, L., Mächler, E., Wüthrich, R. & Altermatt, F. (2020) Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nature Communications*, **11**(1), 1–12.

Carraro, L., Stauffer, J.B. & Altermatt, F. (2021) How to design optimal eDNA sampling strategies for biomonitoring in river networks. *Environmental DNA*, **3**(1), 157–172.

Castello, L. & Macedo, M.N. (2016) Large-scale degradation of Amazonian freshwater ecosystems. *Global Change Biology*, **22**(3), 990–1007.

Copp, G.H., Robert Britton, J., Cucherousset, J., García-Berthou, E., Kirk, R., Peeler, E. et al. (2009) Voracious invader or benign feline? A review of the environmental biology of European catfish *Silurus glanis* in its native and introduced ranges. *Fish and Fisheries*, **10**(3), 252–282.

Cordier, T. (2020) Bacterial communities' taxonomic and functional turnovers both accurately predict marine benthic ecological quality status. *Environmental DNA*, **2**(2), 175–183.

Cordier, T., Lanzen, A., Apotheloz-Perret-Gentil, L., Stoeck, T. & Pawlowski, J. (2019) Embracing environmental genomics and machine learning for routine biomonitoring. *Trends in Microbiology*, **27**(5), 387–397.

Darwall, W., Bremerich, V., De Wever, A., Dell, A.I., Freyhof, J., Gessner, M.O. et al. (2018) The alliance for freshwater life: a global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **28**(4), 1015–1022.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F. et al. (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*, **26**(21), 5872–5895.

Domisch, S., Amatulli, G. & Jetz, W. (2015) Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scientific Data*, **2**(1), 1–13.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G. et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**(1), 27–46.

Dornelas, M., Gotelli, N.J., Shimadzu, H., Moyes, F., Magurran, A.E. & McGill, B.J. (2019) A balance of winners and losers in the Anthropocene. *Ecology Letters*, **22**(5), 847–854.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F. et al. (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, **120**, 25–36.

Esselman, P.C. & Allan, J.D. (2011) Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. *Freshwater Biology*, **56**(1), 71–88.

Farkas, T., Fodor, E., Kitajka, K. & Halver, J. (2001) Response of fish membranes to environmental temperature. *Aquaculture Research*, **32**(8), 645–655.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S. et al. (2007) The shuttle radar topography mission. *Reviews of Geophysics*, **45**(2), 183.

Ferrier, S. (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology*, **51**(2), 331–363.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**(4), 423–425.

Friedman, J.H. (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis*, **38**(4), 367–378.

Gessesse, A.A. & Melesse, A.M. (2019) Temporal relationships between time series CHIRPS-rainfall estimation and eMODIS-NDVI satellite images in Amhara gion, Ethiopia. In: Melesse, A., Abtew, W. & Senay, G. (Eds.) *Extreme hydrology and climate variability*. Ethiopia: Elsevier, pp. 81–92.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M. & Charnov, E.L. (2001) Effects of size and temperature on metabolic rate. *Science*, **293**(5538), 2248–2251.

Goldberg, C.S., Strickler, K.M. & Pilliod, D.S. (2015) Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, **183**, 1–3.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, **202**, 18–27.

Granitto, P.M., Furlanello, C., Biasioli, F. & Gasperi, F. (2006) Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems*, **83**(2), 83–90.

Grooten, M., Almond, R.E. et al. (2018) *Living planet report 2018: aiming higher*. Gland, Switzerland: WWF International.

Growns, I., Gehrke, P.C., Astles, K.L. & Pollard, D.A. (2003) A comparison of fish assemblages associated with different riparian vegetation types in the Hawkesbury-Nepean River system. *Fisheries Management and Ecology*, **10**(4), 209–220.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2–3), 147–186.

Harrington, J.A., Jr., Schiebe, F.R. & Nix, J.F. (1992) Remote sensing of lake Chicot, Arkansas: monitoring suspended sediments, turbidity, and Secchi depth with Landsat MSS data. *Remote Sensing of Environment*, **39**(1), 15–27.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S. et al. (2015) Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, **1**(1), 4–18.

Heino, M., Dieckmann, U. & Godø, O.R. (2002) Estimating reaction norms for age and size at maturation with reconstructed immature size distributions: a new technique illustrated by application to northeast arctic cod. *ICES Journal of Marine Science*, **59**(3), 562–575.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**(2–3), 331–341.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199**(2), 142–152.

Hoyer, M.V., Frazer, T.K., Notestein, S.K. & Canfield, D.E., Jr. (2002) Nutrient, chlorophyll, and water clarity relationships in Florida's nearshore coastal waters with comparisons to freshwater lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**(6), 1024–1031.

Humpl, M. & Pivnička, K. (2006) Fish assemblages as influenced by environmental factors in streams in protected areas of The Czech Republic. *Ecology of Freshwater Fish*, **15**(1), 96–103.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, **27**(3), 151–159.

Kail, J., Brabec, K., Poppe, M. & Januschke, K. (2015) The effect of river restoration on fish, macroinvertebrates and aquatic macrophytes: a meta-analysis. *Ecological Indicators*, **58**, 311–321.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M. & Brumby, S.P. (2021) Global land use/land cover with sentinel 2 and deep learning. In: *2021 IEEE international geoscience and remote sensing symposium IGARSS*. Brussels, Belgium: IEEE, pp. 4704–4707.

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 Biology*, **25**(3), 811–826.

Knudby, A., Brenning, A. & LeDrew, E. (2010) New approaches to modelling fish–habitat relationships. *Ecological Modelling*, **221**(3), 503–511.

Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H.O., Kampel, M. et al. (2019) Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sensing of Environment*, **224**, 104–118.

Lang, N., Jetz, W., Schindler, K. & Wegner, J.D. (2022) A high-resolution canopy height model of the earth. *arXiv*. Preprint arXiv:2204.08322.

Lee, Z., Shang, S., Qi, L., Yan, J. & Lin, G. (2016) A semi-analytical scheme to estimate secchi-disk depth from Landsat-8 measurements. *Remote Sensing of Environment*, **177**, 101–106.

Lelek, A. (1987) *The freshwater fishes of Europe. Vol. 9: Threatened fishes of Europe*. Wiesbaden: Aula-Verlag.

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J. et al. (2018) Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers. *Environmental Science & Technology*, **52**(20), 11708–11719.

Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R News*, **2**(3), 18–22.

Lin, M., Simons, A.L., Harrigan, R.J., Curd, E.E., Schneider, F.D., Ruiz-Ramos, D.V. et al. (2021) Landscape analyses using eDNA metabarcoding and earth observation predict community biodiversity in California. *Ecological Applications*, **31**(6), e02379.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. et al. (2010) Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143**(11), 2647–2657.

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. *Scientific Reports*, **11**(1), 1–14.

Maavara, T., Chen, Q., Van Meter, K., Brown, L.E., Zhang, J., Ni, J. et al. (2020) River dam impacts on biogeochemical cycling. *Nature Reviews Earth & Environment*, **1**(2), 103–116.

Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A. et al. (2020) Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, **7**, 197.

Mann, R. & Blackburn, J. (1991) The biology of the eel *Anguilla anguilla* (L.) in an English chalk stream and interactions with juvenile trout *Salmo trutta* L. and salmon *Salmo salar* L. *Hydrobiologia*, **218**(1), 65–76.

Maridet, L., Wasson, J.-G., Philippe, M., Amoros, C. & Naiman, R. (1998) Trophic structure of three streams with contrasting riparian vegetation and geomorphology. *Archiv für Hydrobiologie*, **144**(1), 61–85.

Markovic, D., Freyhof, J. & Wolter, C. (2012) Where are all the fish: potential of biogeographical maps to project current and future distribution patterns of freshwater species. *PLoS One*, **7**(7), e40530.

McCullagh, P. & Nelder, J.A. (2019) *Generalized linear models*. Oxfordshire, UK: Routledge.

Mills, C. & Mann, R. (1985) Environmentally-induced fluctuations in year-class strength and their implications for management. *Journal of Fish Biology*, **27**, 209–226.

Morgan, B.E., Bolger, D.T., Chipman, J.W. & Dietrich, J.T. (2020) Lateral and longitudinal distribution of riparian vegetation along an ephemeral river in Namibia using remote sensing techniques. *Journal of Arid Environments*, **181**, 104220.

Nagendra, H. (2001) Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, **22**(12), 2377–2400.

Natekin, A. & Knoll, A. (2013) Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, **7**, 21.

Nunn, A., Cowx, I., Frear, P. & Harvey, J. (2003) Is water temperature an adequate predictor of recruitment success in cyprinid fish populations in lowland rivers? *Freshwater Biology*, **48**(4), 579–588.

Olden, J.D. & Jackson, D.A. (2002) A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, **47**(10), 1976–1995.

Olmanson, L.G., Bauer, M.E. & Brezonik, P.L. (2008) A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sensing of Environment*, **112**(11), 4086–4097.

Ozesmi, S.L. & Bauer, M.E. (2002) Satellite remote sensing of wetlands. *Wetlands Ecology & Management*, **10**(5), 381–402.

Page, B.P., Kumar, A. & Mishra, D.R. (2018) A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom. *International Journal of Applied Earth Observation and Geoinformation*, **66**, 69–81.

Palmer, M.A., Reidy Liermann, C.A., Nilsson, C., Flörke, M., Alcamo, J., Lake, P.S. et al. (2008) Climate change and the world's river basins: anticipating management options. *Frontiers in Ecology and the Environment*, **6**(2), 81–89.

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. *Molecular Ecology*, **29**(22), 4258–4264.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A. et al. (2018) The future of biotic indices in the ecogenomic era: integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, **637**, 1295–1310.

Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C. & Guisan, A. (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, **26**(3), 275–287.

Pettorelli, N., Safi, K. & Turner, W. (2014) Satellite remote sensing, biodiversity research and conservation of the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1643), 20130190.

Polanco Fernández, A., Marques, V., Fopp, F., Juhel, J.-B., Borrero-Pérez, G.H., Cheutin, M.-C. et al. (2021) Comparing environmental DNA metabarcoding and underwater visual census to monitor tropical reef fishes. *Environmental DNA*, **3**(1), 142–156.

Pont, D., Logez, M., Carrel, G., Rogers, C. & Haidvogl, G. (2015) Historical change in fish species distribution: shifting reference conditions and global warming effects. *Aquatic Sciences*, **77**(3), 441–453.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A. et al. (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, **8**(1), 1–13.

Raes, N. & ter Steege, H. (2007) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**(5), 727–736.

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S. et al. (2020) Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, **239**, 111626.

Reul, N., Grodsky, S., Arias, M., Boutin, J., Catany, R., Chapron, B. et al. (2020) Sea surface salinity estimates from spaceborne L-band radiometers: an overview of the first decade of observation (2010–2019). *Remote Sensing of Environment*, **242**, 111769.

Revenga, C., Campbell, I., Abell, R., De Villiers, P. & Bryer, M. (2005) Prospects for monitoring freshwater ecosystems towards the 2010 targets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1454), 397–413.

Rowan, G.S. & Kalacska, M. (2021) A review of remote sensing of submerged aquatic vegetation for non-specialists. *Remote Sensing*, **13**(4), 623.

Sandström, A., Philipson, P., Asp, A., Axenrot, T., Kinnerbäck, A., Ragnarsson-Stabo, H. et al. (2016) Assessing the potential of remote sensing-derived water quality data to explain variations in fish assemblages and to support fish status assessments in large lakes. *Hydrobiologia*, **780**(1), 71–84.

Schwager, P. & Berg, C. (2021) Remote sensing variables improve species distribution models for alpine plant species. *Basic and Applied Ecology*, **54**, 1–13.

Shuter, B. & Post, J. (1990) Climate, population viability, and the zoogeography of temperate fishes. *Transactions of the American Fisheries Society*, **119**(2), 314–336.

Siefert, R.E., Spoor, W.A. & Syrett, R.F. (1973) Effects of reduced oxygen concentrations on northern pike (*Esox lucius*) embryos and larvae. *Journal of the Fisheries Board of Canada*, **30**(6), 849–852.

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Flanagan, L.B. et al. (2006) On the use of Modis EVI to assess gross primary productivity of North American ecosystems. *Journal of Geophysical Research: Biogeosciences*, **111**, G04015.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**(8), 2045–2050.

Thomsen, P.F. & Willerslev, E. (2015) Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, 4–18.

Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B. et al. (2023) *biomod2: ensemble platform for species distribution modeling*. R package version 4.2-3.

Turak, E., Harrison, I., Dudgeon, D., Abell, R., Bush, A., Darwall, W. et al. (2017) Essential biodiversity variables for measuring change in global freshwater biodiversity. *Biological Conservation*, **213**, 272–279.

Turner, C.R., Barnes, M.A., Xu, C.C., Jones, S.E., Jerde, C.L. & Lodge, D.M. (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, **5**(7), 676–684.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. & Steininger, M. (2003) Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, **18**(6), 306–314.

van Proosdij, A.S., Sosef, M.S., Wieringa, J.J. & Raes, N. (2016) Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, **39**(6), 542–552.

Vanhellemont, Q. (2020) Automated water surface temperature retrieval from Landsat 8/TIRS. *Remote Sensing of Environment*, **237**, 111518.

Vezza, P., Muñoz-Mas, R., Martinez-Capel, F. & Mouton, A. (2015) Random forests to evaluate biotic interactions in fish distribution models. *Environmental Modelling & Software*, **67**, 173–183.

Vihervaara, P., Auvinen, A.-P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S. et al. (2017) How essential biodiversity variables and remote sensing can help national biodiversity monitoring. *Global Ecology & Conservation*, **10**, 43–59.

Wang, R. & Gamon, J.A. (2019) Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, **231**, 111218.

Watanabe, F.S.Y., Alcantara, E., Rodrigues, T.W.P., Imai, N.N., Barbosa, C.C.F. & da Silva Rotta, L.H. (2015) Estimation of chlorophyll-a concentration and the trophic state of the Barra Bonita hydroelectric reservoir using OLI/Landsat-8 images. *International Journal of Environmental Research and Public Health*, **12**(9), 10391–10417.

Wolter, C. (2007) Temperature influence on the fish assemblage structure in a large lowland river, the lower Oder River, Germany. *Ecology of Freshwater Fish*, **16**(4), 493–503.

Wulder, M.A., LeDrew, E.F., Franklin, S.E. & Lavigne, M.B. (1998) Aerial image texture information in the estimation of northern deciduous and mixed wood forest Leaf Area Index (LAI). *Remote Sensing of Environment*, **64**(1), 64–76.

Yamasaki, E., Altermatt, F., Cavender-Bares, J., Schuman, M.C., Zuppinger-Dingley, D., Garonna, I. et al. (2017) Genomics meets remote sensing in global change studies: monitoring and predicting phenology, evolution and biodiversity. *Current Opinion in Environmental Sustainability*, **29**, 177–186.

Zarkami, R. (2008) *Habitat suitability modelling of pike (Esox lucius) in rivers*. Ghent, Belgium: Ghent University.

Zhang, H., Mächler, E., Morsdorf, F., Niklaus, P.A., Schaepman, M.E. & Altermatt, F. (2023) A spatial fingerprint of land-water linkage of biodiversity uncovered by remote sensing and environmental DNA. *Science of the Total Environment*, **161365**, 161365.

Zhang, H., Yoshizawa, S., Iwasaki, W. & Xian, W. (2019) Seasonal fish assemblage structure using environmental DNA in the Yangtze estuary and its adjacent waters. *Frontiers in Marine Science*, **6**, 515.

Zhang, L., Yang, J., Zhang, Y., Shi, J., Yu, H. & Zhang, X. (2022) eDNA biomonitoring revealed the ecological effects of water diversion projects between Yangtze River and Tai Lake. *Water Research*, **210**, 117994.

Zheng, G. & Moskal, L.M. (2009) Retrieving Leaf Area Index (LAI) using remote sensing: theories, methods and sensors. *Sensors*, **9**(4), 2719–2745.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2020) Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **47**(1), 101–113.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. List of species used in the species distribution models and number of occurrences of the species in the Rhone river.

Figure S1. Diagram of the upstream-directed buffers. Blue circles from light to dark denote buffer size, from local to 20 km upstream of the sampling sites. For better visualisation, the buffer sizes are not drawn to scale.

Figure S2. True skill statistic (TSS) of all the species distribution models (SDMs; 3 types), across 29 species (77 SDMs total).

Figure S3. Analysis of variance (ANOVA) results of the median effective true skill statistic (ΔTSS) across the two buffer types (centred around sampling point and upstream-directed) and the four buffer sizes (5, 10, 15, 20 km).

Figure S4. Area under the ROC curve (AUC) of all the species distribution models (SDMs; 3 types), across 29 species (77 SDMs total).

Figure S5. Cohen's kappa of all the species distribution models (SDMs; 3 types), across 29 species (77 SDMs total).

Figure S6. True skill statistic (TSS) comparison between null models and species distribution models (SDMs).

Figure S7. Area under the ROC curve (AUC) comparison between null models and species distribution models (SDMs).

Figure S8. Cohen's kappa comparison between null models and species distribution models (SDMs).

Figure S9. Effective model performance, along with the number of occurrences of species.

Figure S10. Model uncertainty map.

Figure S11. Variable importance in all the models (top 12 most important variables).

Figure S12. River water surface temperature validation with water temperature measurement stations in Switzerland (Geneva) and France (ST-VALLIER, PIERRELATTE, and ROQUEMAURE).