Earth Syst. Sci. Data, 18, 1541–1559, 2026
https://doi.org/10.5194/essd-18-1541-2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Open Access Earth System Science Data

Environment90m – globally standardized environmental variables for freshwater science at high spatial resolution

Jaime R. García-Márquez[1], Afroditi Grigoropoulou[1], Thomas Tomiczek[1], Marlene Schürz[1,2,3], Vanessa Bremerich[1], Yusdiel Torres-Cambas[1], Merret Buurman[1], Kristi Bego[1], Giuseppe Amatulli[4,5], and Sami Domisch[1]

[1]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Department of Community and Ecosystem Ecology, Müggelseedamm 310, 12587 Berlin, Germany
[2]Freie Universität Berlin, Department of Biology, Chemistry, Pharmacy, Institute of Biology, Königin-Luise-Str. 1–3, 14195 Berlin, Germany
[3]Universität für Bodenkultur Wien, Institute of Hydrobiology and Aquatic Ecosystem Management, Gregor-Mendel-Straße 33, 1180 Vienna, Austria
[4]Yale University, School of the Environment, 195 Prospect Street, New Haven, CT, 06511, USA
[5]Spatial Ecology, 35A, Hazlemere Road, Penn, Buckinghamshire, HP10 8AD, United Kingdom

Correspondence: Jaime R. García-Márquez (jaime.marquez@igb-berlin.de) and Sami Domisch (sami.domisch@igb-berlin.de)

Received: 11 July 2025 – Discussion started: 5 August 2025
Revised: 16 January 2026 – Accepted: 16 January 2026 – Published: 27 February 2026

Abstract. The current loss of freshwater biodiversity calls for immediate action, including the mobilisation of existing data and tools to support long-term conservation. Yet, establishing a global baseline for the spatial distribution of freshwater habitats and the biodiversity they host remains difficult. Such task would require standardized, high-resolution environmental information to characterise freshwater habitats anywhere in the world. To address this challenge, we present the Environment90m dataset, which aggregates a large number of environmental layers into each of the 726 million sub-catchments of the Hydrography90m dataset, corresponding to single stream segments. Specifically, Environment90m includes 45 variables related to topography and hydrography, 19 climate variables for the observation period of 1981–2010, as well as projections for 2041–2070 and 2071–2100 under the Shared Socioeconomic Pathways (SSPs) 1.26, 3.70 and 5.85, and three global circulation models (UKESM1-0-LL, MPI-ESM1-2-HR and IPSL-CM6A-LR). Moreover, Environment90m includes 22 land cover categories for the annual time-series data from 1992–2020. In addition, we provide 15 soil variables and information on aridity and modelled streamflow. Summary statistics (i.e., mean, min, max, range, sd) are provided for all continuous variables, while for categorical data, the proportion of each category is calculated within each of the sub-catchments. The data is available at https://hydrography.org/environment90m (last access: 4 February 2026). To facilitate data download and processing, we provide dedicated functions within the hydrographr R-package, and extend these also to new functions for processing upstream data of lakes. For all underlying calculations, we used the open-source tools GDAL/OGR, GRASS-GIS and AWK, so that custom data can be easily generated using the hydrographr R-package. Environment90m, along with the tools, provides an array of opportunities for research and application in spatial freshwater biodiversity science, specifically biogeographic analyses and conservation exercises in freshwater ecosystems. The metadata of the Environment90m dataset is stored at https://doi.org/10.18728/igb-fred-995.0 (García Márquez et al., 2025a).

Published by Copernicus Publications.

1 Introduction

Freshwater ecosystems and biodiversity are considered to be more threatened than their terrestrial and marine counterparts (WWF, 2020; Tickner et al., 2020). Advances towards the protection of freshwater habitats remain elusive, despite recent efforts to secure their long-term protection. Examples include the so-called "30-by-30" protection target, which aims to protect 30 % of Earth's lands, oceans, coastal areas and inland waters (The Post-2020 Global Biodiversity Framework, Hughes, 2023). In addition, the recent EU Nature Restoration Law also aims to restore river connectivity (Stoffers et al., 2024). In the context of these political initiatives, large-scale standardized analyses on the spatial, environmental and biotic characterization of freshwater habitats and their connectivity are required to, e.g., answer the question: "which areas should be prioritized for protection?". Addressing this question requires at minimum a detailed baseline of the present-day spatial distribution of the environmental characteristics of freshwater habitats. Only after establishing a baseline, the environmental changes, or changes in biodiversity can be measured and quantified. Such knowledge allows to assess freshwater biodiversity using, e.g., species distribution modelling techniques (Bellin et al., 2022). Likewise, it allows to also perform connectivity-related analyses to support the restoration of fragmented rivers (Hermoso, 2025).

In the freshwater realm, information on such environmental characteristics should ideally be available at very high spatial resolution. This is because (i) it can be attributed to the corresponding water body, e.g., a specific stream segment and (ii) environmental characteristics are not aggregated across large areas, avoiding issues related to the Modifiable Area Unit Problem (MAUP) (Jelinski and Wu, 1996). The MAUP is a statistical feature which occurs when data is aggregated to spatial units, whose size may influence the aggregated values (e.g., by using grid cells of varying size). In the freshwater realm, spatial units are more realistic when they correspond to drainages, larger sub-catchments or standing water bodies such as lakes. Therefore environmental information is commonly aggregated within these units. A key goal is thus to have environmental information at the highest possible spatial resolution, allowing to incorporate the crucial feature of longitudinal connectivity, while still achieving computational efficiency of workflows. The environmental conditions along the dendritic network structure can be depicted following various concepts: the River Continuum Concept (Vannote et al., 1980), macrosystem theory (Thorp, 2014) or functional process zones (Maasri et al., 2019). At the same time, tributary inputs, lateral connectivity with floodplains, and discontinuities caused by natural or anthropogenic disturbances also play a role in shaping the environmental conditions along the dendritic stream network (Ward and Stanford, 1983, 1995; Benda et al., 2004). To characterize this environmental continuum along the network requires, in turn, to pinpoint the relevant environmental conditions and processes to single network segments. Hence, the spatial aggregation of environmental information, which usually comes in gridded datasets at, e.g., 1 km spatial resolution, has to match the spatial configuration of the water bodies (Brunner et al., 2024; Friedrichs-Manthey et al., 2020). In this regard, sub-catchments which correspond to single stream segments are, unlike pixels, non-randomly distributed across the surface and follow the topographical and topological gradients in the landscape (Brunner et al., 2024).

Therefore, in freshwater systems, the sub-catchments of single stream segments can be considered as the smallest hydrological units. They have their own ecological and hydrological boundaries, encompassing the aquatic-terrestrial linkages (Linke et al., 2007). They also feature the same connectivity as the network, but also allow including the terrestrial landscape into the analysis workflow, which is of interest when performing biogeographic analyses of semi-aquatic organisms, such as aquatic insects, amphibians or mammals, which rely on both ecosystems.

Aggregating environmental variables across sub-catchments is often done on a case-by-case basis, while there are only few datasets that cover wide spatial gradients. A global extent is given by the HydroATLAS (Linke et al., 2019) and the HydroLAKES datasets (Messager et al., 2016; Lehner et al., 2022), as well as a near-global aggregation of upstream-catchment variables (Domisch et al., 2015) which follow the HydroSHEDS river network (Lehner et al., 2008). National efforts include the StreamCAT Dataset (Hill et al., 2016) and LakeCAT Dataset (Hill et al., 2018) which correspond to the high-resolution NHDPlusV2 river network in the United States (McKay et al., 2012).

In contrast to all previous examples, the Hydrography90m dataset (Amatulli et al., 2022), includes the most detailed stream network channel at 90 m spatial resolution. This global network incorporates for the first time headwater stream segments, that is, streams of 1st and 2nd Strahler order (Strahler, 1957). Although some estimates indicate that 70 % of global stream networks consist of 1st and 2nd order streams (Lowe and Likens, 2005; Leopold et al., 2020), they have been neglected so far in regional and global biodiversity assessments. Hydrography90m comprises 1.6 million drainage basins, 726 million stream segments and sub-catchments, along with 42 stream-topographical and -topological variables. Each sub-catchment corresponds to one stream segment, sharing a unique ID.

With this level of detail, aggregating environmental variables across sub-catchments implies computationally heavy calculations and data storage challenges. With the introduction of the Environment90m dataset, we expect to facilitate globally standardized analyses and modelling workflows by aggregating environmental variables at global scale and at high spatial resolution based on the Hydrography90m stream network (Amatulli et al., 2022). Environment90m includes 45 variables related to topography and hydrography, 19 cli-

mate variables for the observation period of 1981–2010, as well as projections for 2041–2070 and 2071–2100 under the Shared Socioeconomic Pathways (SSPs) 1.26, 3.70 and 5.85, and three global circulation models (UKESM, MPI and IPSL). Moreover, Environment90m includes 22 land cover categories for the annual time-series data from 1992 to 2020. In addition, we provide 15 soil variables and information on aridity and modelled stream flow. Summary statistics (i.e., mean, minimum, maximum, range, standard deviation) are provided for all continuous variables. For categorical variables, the proportion of each category is calculated within each sub-catchment. The current version of the Environment90m dataset is 4.1 terabytes (TB) in size. The data is available at https://hydrography.org/environment90m (last access: 4 February 2026) (see also Sect. 2.3).

To mobilize the Environment90m data integration into workflows, we provide two options: first, we have developed custom functions for batch downloading, processing and integrating the data with the Hydrography90m network, as well as for performing custom workflows, within the hydrographr R-package (Schürz et al., 2023). Second, all Environment90m aggregated variables are also available on the GeoFRESH online platform (available at https://geofresh.org/, last access: 4 February 2026, Domisch et al., 2024), which allows to quickly download the variables for point locations globally, both for the given sub-catchments where the points are located and their entire upstream catchments. Moreover, a new set of hydrographr functions allows to aggregate the data over lakes, as well as single river-lake intersections (Tomiczek et al., 2024). For example, users can retrieve the Environment90m information for each lake's upstream catchment area. We showcase the workflow in this exemplary lake vignette, available at https://glowabio.github.io/hydrographr/articles/case_study_lakes.html (last access: 4 February 2026).

2 Methods

2.1 Workflow for the development of the Environment90m dataset

The Environment90m dataset consists of tabular data including the summary statistics of 104 environmental variables from 7 different datasets, which we aggregated over each of the 726 million unique sub-catchments of the Hydrography90m dataset (Amatulli et al., 2022), yielding different summary statistics. The output tables are available in different formats (Fig. 1, and see also Sect. 2.3). We selected variables that are available continuously (i.e., range-wide) at the global scale with a resolution of 1 km or less, and which are widely used or can potentially be used to describe freshwater habitats and biodiversity distribution patterns. The following sections describe the original datasets and the processes we followed for their aggregation.

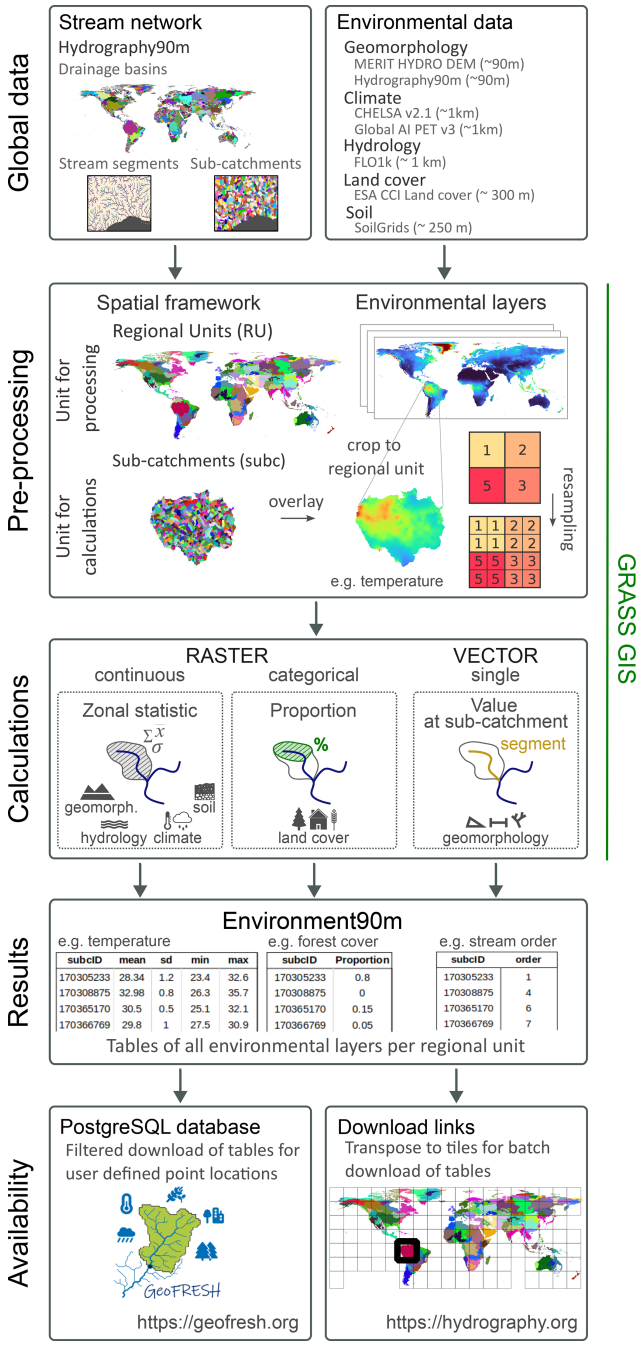


Figure 1. Workflow for the development of the Environment90m dataset. The process starts with the selection of environmental variables and their resampling to match the 90 m resolution of the Hydrography90m sub-catchment layer. The aggregation is done through several raster and vector calculations. The final tables are available to download at https://hydrography.org/environment90m/ (last access: 4 February 2026) and can be also queried in the GeoFRESH online platform at https://geofresh.org/ (last access: 4 February 2026).

We started the procedure by creating a working session in GRASS GIS (Neteler et al., 2012) for each of the 166 Re-

gional Units (RUs) defined in Amatulli et al. (2022) (see also Fig. 7 therein). Regional units are groups of one large or several entire drainage basins, ensuring that the whole area of the sub-catchments belonging to a drainage basin is included in the RU. This design aims to improve the efficiency of computational calculations. The GRASS GIS session initialization was made with the raster file of each RU containing the sub-catchments, which automatically provided the geographical extent and the resolution (90 m) of the raster file as the default settings of the session. The default coordinate reference system was set to the World Geodetic System 1984 (WGS 84) with coordinates expressed as latitude and longitude, as defined by the `EPSG:4236`.

The raster files of the environmental variables were then read into each of the GRASS GIS sessions, where they were cropped and resampled to the same extent and resolution using the default settings. In all our cases, the original raster files had a resolution equal or lower than 90 m. In case the environmental layers had a lower resolution, as CHELSA climate at a 30 arcsec (1 km) resolution, their grid cells were resampled to 90 m without interpolation, i.e., all new 90 m cells were assigned the same value as 1 km cells if they overlapped (GRASS Development Team, 2024). This procedure is explained in Fig. 2.

We selected between three possible methods to calculate summary statistics when creating the output tables, following the properties of each environmental dataset:

1. *Zonal statistics.* Calculation of the mean, standard deviation, range, minimum, and maximum of the environmental layer within each sub-catchment. Environmental layers with continuous values (e.g., temperature) fell into this category. The calculations were done within the GRASS GIS environment using the `r.univar` function.

2. *Proportion.* The proportion of the variable (i.e., variable categories) within each sub-catchment. Categorical data, namely land cover, belonged to this category. The calculations were done within the GRASS GIS environment by dividing the number of pixels of each land cover category in each sub-catchment by the total number of pixels in the sub-catchment.

3. *Value at sub-catchment.* Hydrography90m provides a vector file with a list of attributes for every stream segment globally. Since every sub-catchment shares a unique ID with each stream segment, the value assigned to the sub-catchment corresponds to the value of the different attributes in the stream segment vector file (Amatulli et al., 2022). Examples of these attributes include stream length or Strahler stream order (Strahler, 1957).

All calculations were performed in parallel using the High Performance Computing (HPC) facility at Yale University.

2.2 Description of the underlying environmental datasets

The following sections describe the underlying environmental datasets used to generate Environment90m. The tables refer to the properties of the original datasets. The exact abbreviations used in Tables 1–7 should be used to retrieve the variables from the Environment90m dataset.

2.2.1 Stream network data

The Hydrography90m is a high-resolution (∼ 90 m) dataset delineating a global stream channel network (Amatulli et al., 2022) and serves as the basis of Environment90m. The calculation of Hydrography90m used the MERIT Hydro Digital Elevation Model at 3 arcsec (∼ 90 m at the Equator) (Yamazaki et al., 2017). The main feature of Hydrography90m is the delineation of small headwater streams. In addition, this dataset includes a number of stream topographic and topological properties (see Table 1), which have already been used for freshwater habitat analyses (e.g., Schürz et al., 2025).

2.2.2 Climate

We derived high-resolution climate variables from the CHELSA v2.1 dataset available at "https://chelsa-climate.org/ (last access: 4 February 2026)" (Karger et al., 2017, 2021). We used 19 bioclimatic variables (bio 1 to 19) at 30 arcsec (ca. 1 km) resolution (Table 2). We aggregated the reference period available as a long-term annual average (1981–2010). Data for future climate projections (i.e., periods 2041–2070 and 2071-2100) were aggregated by selecting individual pairings of three global circulation models (GCMs, i.e., mpi-esm1-2-hr, ukesm1-0-ll, ipsl-cm6a-lr) and three Shared Socioeconomic Pathways (SSPs, i.e., SSP1-RCP2.6 [Sustainability – lowest emission scenario], SSP3-RCP7 [Regional Rivalry – middle emission scenario], and SSP5-RCP8.5 [Fossil Fuel Development – highest emission scenario]; Ebi et al., 2014; O'Neill et al., 2017). With this selection we cover short and long term future projections, and a wide spectrum of potential future trajectories given social, technological, economical and environmental changes. This diversity of options should facilitate the evaluation of uncertainties in climate projection exercises. Since its publication, the CHELSA v2.1 dataset has been used in studies on the suitability of Natura 2000 freshwater habitats for threatened species (e.g., Basen et al., 2022). It has also been used to track the niche and distribution dynamics of invasive freshwater fauna (e.g., Guareschi et al., 2024).

2.2.3 Land cover

For land cover data, we used the consistent global land cover maps of the Land Cover European Space Agency (ESA) Climate Change Initiative (CCI) project (ESA, 2017). We cal-

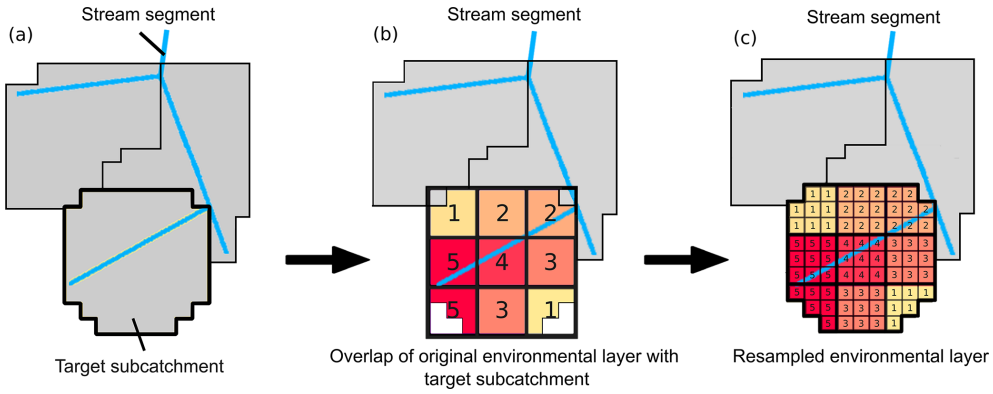


Figure 2. Resampling procedure (from lower to higher resolution) when reading environmental datasets into the default settings of a GRASS GIS session. **(a)** shows an example of different stream segments and their corresponding sub-catchments. The sub-catchment with the bold boundary is the target sub-catchment to illustrate the resampling procedure. **(b)** shows the overlap of the environmental data (originally with a lower resolution, e.g., 1 km) with the target sub-catchment. In **(c)** the original dataset is resampled to the 90 m resolution of the target sub-catchment, such that all new 90 m raster cells have the corresponding value as the lower resolution cell.

culated the proportion of each land cover within the sub-catchments. We used level 1 land cover categories, which consists of 22 categories (Table 3) developed to describe land cover patterns globally. The 22 categories were selected based on their consistency in global coverage over the entire time period outlined in the Land Cover CCI product user guide (ESA, 2017). The annual data are available for the years 1992 to 2020. The ESA CCI land cover data have been useful in research that aims to identify patterns and hotspots of global land cover transitions (Liu et al., 2018). Other researchers have used the dataset to e.g. enhance water balance analyses in tropical regions (Tan et al., 2021).

2.2.4 Soil

The soil variables were sourced from the global gridded soil information dataset, SoilGrids250 v2.0 (Hengl et al., 2017). This dataset represents global chemical and physical soil properties (Table 4). Each of the variables was originally provided at six standard depths (with the exception of depth to bedrock and soil organic carbon content) and at a spatial resolution of 250 m. To integrate all available soil depths (up to 200 cm), we calculated the weighted average for each soil property originally measured at different depths, following the GlobalSoilMap specifications described in Arrouays et al. (2014).

2.2.5 Elevation

To represent elevation we used the 90 m resolution Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT DEM) (Yamazaki et al., 2017) (Table 5). The error removal procedures applied to this dataset have improved its vertical accuracy. This dataset was also used as the basis for the creation of the Hydrography90m dataset (Amatulli et al., 2022). Other research has made use of the MERIT DEM to enable machine learning-based estimations of dynamic water depth and volume of global lakes (Lv et al., 2025), or to conduct and subsequently evaluate continental-scale river hydrodynamic modelling (Modi et al., 2022).

2.2.6 Stream flow

For stream (water) flow we used the FLO1K dataset which comprises the modelled mean, maximum and minimum annual flow for each year in the period 1960–2015, provided as spatially continuous gridded layers at 30 arcsec (ca. 1 km) (Barbarossa et al., 2018) (Table 6). For Environment90m, we only used the data from 1980–2010 and averaged the data across this time frame, to match the CHELSA observed climate dataset (Sect. 2.2.2).

2.2.7 Global Aridity Index and Potential Evapotranspiration

This dataset provides high-resolution (30 arcsec, ca. 1 km) global raster data on evapotranspiration patterns and rainfall deficit for potential vegetation growth. Global Aridity and Potential Evapotranspiration are both modeled using data available from WorldClim Global Climate Data. The data is available for the 1970–2000 period (Zomer and Trabucco, 2022) (Table 7). Since its publication, the Global Aridity and Potential Evapotranspiration dataset has been used e.g. in data-driven modelling of freshwater ecosystems in Europe (Almeida and Cabral, 2023), as well as for the development of a global data set on drainage systems (He et al., 2024).

2.3 Accessing the Environment90m dataset

The initial set of tables contained the environmental information for all sub-catchments within each RU (see

Table 1. List of variables derived from the Hydrography90m dataset. The original values of each variable are assigned on a 90m raster cell basis. A *focal cell* refers to the target cell or pixel where the calculation of the variables is being done. The term *init* refers to the initial node of a stream segment, i.e., the most upstream cell. The scale factor attached to the units column should be applied to calculate the real values.

Hydrography90m (Amatulli et al., 2022)				
spatial resolution	90 m			
temporal resolution	–			
time range	–			
statistics	mean, standard deviation, minimum, maximum, range. for stream order and stream reach variables: value of stream segment			
Variable Type	Variable	Abbreviation	Unit	Description
Flow accumulation	Flow accumulation	accumulation	km^2	Accumulated number of cells that drain through each cell
Stream slope	Cell maximum curvature	slope_curv_max_dw_cel	m^{-1} (Scale factor 10^6)	Maximum linear curvature along watercourse, calculated between highest upstream cell, focal cell and downstream cell
	Cell minimum curvature	slope_curv_min_dw_cel	m^{-1} (Scale factor 10^6)	Minimum linear curvature along watercourse, calculated between lowest upstream cell, focal cell and downstream cell
	Cell elevation difference	slope_elv_dw_cel	m	Difference between elevation of focal cell and downstream cell
	Cell gradient	slope_grad_dw_cel	(Scale factor 10^6)	Downstream elevation difference divided by distance
Stream distance	Shortest distance to drainage divide	stream_dist_up_near	m	Shortest upstream distance between focal cell and the nearest sub-catchment drainage divide
	Longest distance to drainage divide	stream_dist_up_farth	m	Longest upstream distance between focal cell and the nearest sub-catchment drainage divide
	Nearest downstream stream cell	stream_dist_dw_near	m	Distance between focal cell and its nearest downstream stream cell
	Outlet cell in the network	outlet_dist_dw_basin	m	Distance between focal cell and the outlet cell in the network
	Downstream stream node cell	outlet_dist_dw_scatch	m	Distance between focal cell and the downstream stream node cell
	Euclidean distance	stream_dist_proximity	m	Euclidean distance between focal cell and the stream network
Elevation	Shortest path	stream_diff_up_near	m	Elevation difference of the shortest path from focal cell to the sub-catchment drainage divide
	Longest path	stream_diff_up_farth	m	Elevation difference of the longest path from focal cell to the sub-catchment drainage divide
	Nearest downstream stream grid cell	stream_diff_dw_farth	m	Elevation difference between focal cell and its nearest downstream stream grid cell
	Outlet grid cell in the network	outlet_diff_dw_basin	m	Elevation difference between focal cell and the outlet grid cell in the network
	Downstream stream node cell	outlet_diff_dw_scatch	m	Elevation difference between focal cell and the downstream stream node cell
Segment properties	Segment downstream mean gradient	channel_grad_dw_seg		Segment downstream mean gradient (between focal cell and the node/outlet)
	Segment upstream mean gradient	channel_grad_up_seg		Segment upstream mean gradient (between focal cell and the init/node)
	Cell upstream gradient	channel_grad_up_cel		Cell upstream gradient (between focal cell and next cell)
	Cell stream course curvature	channel_curv_cel		Cell stream course curvature (focal cell)

Table 1. Continued.

Variable Type	Variable	Abbreviation	Unit	Description
	Segment downstream elevation difference	channel_elv_dw_seg		Segment downstream elevation difference (between focal cell and the node/outlet)
	Segment upstream elevation difference	channel_elv_up_seg		Segment upstream elevation difference (between focal cell and the init/node)
	Cell upstream elevation difference	channel_elv_up_cel		Cell upstream elevation difference (between focal cell and next cell)
	Cell downstream elevation difference	channel_elv_dw_cel		Cell downstream elevation difference (between focal cell and next cell)
	Segment downstream distance	channel_dist_dw_seg		Segment downstream distance (between focal cell and the node/outlet)
	Segment upstream distance	channel_dist_up_seg		Segment upstream distance (between focal cell and the init/node)
	Cell upstream distance	channel_dist_up_cel		Cell upstream distance (between focal cell and next cell)
Stream order	Strahler's stream order	order_strahler		
	Shreve's stream magnitude	order_shreve		
	Horton's stream order	order_horton		
	Hack's stream order	order_hack		
	Topological dimension of streams	order_topo		
Stream reach	Length of the stream reach	length	m	Length of the stream reach
	Straight length	stright	m	Length of the stream as straight line
	Sinusoid of the stream reach	sinosoid		Fractal dimension: stream length/straight stream length
	Accumulated length	cum_length	m	Length of stream from source
	Distance to outlet	out_dist	m	Distance of current stream init from outlet
	Source elevation	source_elev	m	Elevation of stream init
	Outlet elevation	outlet_elev	m	Elevation of stream outlet
	Elevation drop	elev_drop	m	Difference between source_elev and outlet_elev + drop outlet
	Outlet drop	out_drop	m	Drop at the outlet of the stream
	Gradient	gradient	m	Mean gradient of the sub-catchment (downstream elevation difference divided by distance)
Flow index	Stream power index	spi		Measure of the erosive power of flowing water (Moore et al., 1991)
	Sediment transportation index	sti		Metric describing the erosion and deposition of sediments (Mojaddadi et al., 2017)
	Compound topographic index	cti		A steady state wetness index, also known as topographic wetness index (TWI) (Beven and Kirkby, 1979)
Stream connectivity	Connectivity	connections		Attribute table with the sub-catchment id of the next stream segment (downstream: *next_stream*), and two or more contributing streams (upstream: *prev_stream*)

Table 2. List of variables derived from the CHELSA dataset. The scale and offset values provided in the description of some variables are factors used to convert the raw values to meaningful values, for example from raw data to Celcius degree for the temperature related variables. The equation to apply the factors is Corrected Value = (Raw Value × Scale) + Offset.

Climatologies at high resolution for the earth's land surface areas. CHELSA v2.1 (Karger et al., 2017, 2021)				
spatial resolution	1 km			
temporal resolution	Long Term (30 years) Annual Average			
time range	1981–2010, 2041–2070, 2071–2100			
circulation models	ipsl-cm6a-lr, mpi-esm1-2-hr, ukesm1-0-ll			
shared socioeconomic pathways	ssp126, ssp370, spp586			
statistics	mean, standard deviation, minimum, maximum, range			
Variable Type	Variable	Abbreviation	Unit	Description
Temperature	Annual mean temperature	bio01	°C	Scale = 0.1, Offset = −273.15: Mean annual daily mean air temperatures averaged over 1 year
	Mean diurnal range	bio02	°C	Scale = 0.1: Mean diurnal range of temperatures averaged over 1 year
	Isothermality	bio03	°C	Scale = 0.1: Ratio of diurnal variation to annual variation in temperatures
	Temperature seasonality	bio04	°C/100	Scale = 0.1: Standard deviation of the monthly mean temperatures
	Max temperature of warmest month	bio05	°C	Scale = 0.1, Offset = −273.15: The highest temperature of any monthly daily mean maximum temperature
	Min temperature of coldest month	bio06	°C	Scale = 0.1, Offset = −273.15: The lowest temperature of any monthly daily mean minimum temperature
	Temperature annual range	bio07	°C	Scale = 0.1: The difference between the Maximum Temperature of Warmest month and the Minimum Temperature of Coldest month
	Mean temperature of wettest quarter	bio08	°C	Scale = 0.1, Offset = −273.15: The wettest quarter of the year is determined (to the nearest month)
	Mean temperature of driest quarter	bio09	°C	Scale = 0.1, Offset = −273.15: The driest quarter of the year is determined (to the nearest month)
	Mean Temperature of warmest Quarter	bio10	°C	Scale = 0.1, Offset = −273.15: The warmest quarter of the year is determined (to the nearest month)
	Mean Temperature of coldest Quarter	bio11	°C	Scale = 0.1, Offset = −273.15: The coldest quarter of the year is determined (to the nearest month)
Precipitation	annual precipitation	bio12	mm	Scale = 0.1: Accumulated precipitation amount over 1 year
	Precipitation of wettest month	bio13	mm	Scale = 0.1: The precipitation amount of the wettest month
	Precipitation of driest month	bio14	mm	Scale = 0.1: The precipitation amount of the driest month
	Precipitation seasonality	bio15		Scale = 0.1: The Coefficient of Variation is the standard deviation of the monthly precipitation estimates expressed as a percentage of the mean of those estimates (i.e. the annual mean)
	Precipitation of wettest quarter	bio16	mm	Scale = 0.1: The wettest quarter of the year is determined (to the nearest month)
	Precipitation of driest quarter	bio17	mm	Scale = 0.1: The driest quarter of the year is determined (to the nearest month)
	Precipitation of warmest quarter	bio18	mm	Scale = 0.1: The warmest quarter of the year is determined (to the nearest month)
	Precipitation of coldest quarter	bio19	mm	Scale = 0.1: The coldest quarter of the year is determined (to the nearest month)

Table 3. List of variables (i.e., land cover categories) derived from the ESA land cover maps. The numbers within parentheses in the description make reference to the coding of the land cover categories at level 2.

Consistent global land cover maps: ESA CCI land cover (ESA, 2017)				
spatial resolution	300 m			
temporal resolution	Annual			
time range	1992–2020			
statistics	Proportion of each land cover in the sub-catchment			
Variable Type	Variable	Abbreviation	Unit	Description
Land Cover	Cropland	c10	presence	Cropland, rainfed (10, 11, 12)
	Cropland	c20	presence	Cropland, irrigated or post-flooding (20)
	Cropland/natural vegetation	c30	presence	Mosaic cropland (> 50 %) – natural vegetation (tree, shrub, herbaceous cover) (< 50 %) (30)
	Natural vegetation/cropland	c40	presence	Mosaic natural vegetation (tree, shrub, herbaceous cover) (> 50 %)/cropland (< 50 %) (40)
	Tree cover, broadleaved, evergreen	c50	presence	Tree cover, broadleaved, evergreen, closed to open (> 15 %) (50)
	Tree cover, broadleaved, deciduous	c60	presence	Tree cover, broadleaved, deciduos, closed to open (> 15 %) (60 61 62)
	Tree cover, needleleaved, evergreen	c70	presence	Tree cover, needleleaved, evergreen, closed to open (> 15 %) (70 71 72)
	Tree cover, needleleaved, deciduous	c80	presence	Tree cover, needleleaved, deciduous, closed to open (> 15 %) (80 81 82)
	Tree cover, mixed leaf type	c90	presence	Tree cover, mixed leaf type (broadleaved and needleleaved) (90)
	Tree and shrub	c100	presence	Mosaic tree and shrub (> 50 %)/herbaceous cover (< 50 %) (100)
	Herbaceous/tree and shrub	c110	presence	Mosaic herbaceous cover (> 50 %)/tree and shrub (< 50 %) 110)
	Shrubland	c120	presence	Shrubland (120 121 122)
	Grassland	c130	presence	Grassland (130)
	Lichens, mosses	c140	presence	Lichens, mosses (140)
	Sparse vegetation	c150	presence	Sparse vegetation (tree, shrub, herbaceous cover) (< 15 %) (150, 151, 152, 153)
	Tree cover, flooded, fresh/brackish water	c160	presence	Tree cover, flooded, fresh or brackish water (160)
	Tree cover, flooded, saline water	c170	presence	Tree cover, flooded, saline water (170)
	Shrub or herbaceous	c180	presence	Shrub or herbaceous cover, flooded, fresh – saline – brackish water (180)
	Urban areas	c190	presence	Urban areas (190)
	Bare areas	c200	presence	Bare areas (200 201 202)
	Water bodies	c210	presence	Water bodies (210)
	Snow and ice	c220	presence	Permanent snow and ice (220)

Table 4. List of variables derived from the SOILGRID database.

SoilGrids: global gridded soil information (Hengl et al., 2017)				
spatial resolution	250 m			
temporal resolution	–			
time range	–			
statistics	mean, standard deviation, minimum, maximum, range			
Variable Type	Variable	Abbreviation	Unit	Description
Soil	Derived saturated water content	awcts	–	
	Clay content	clyppt	%	
	Sand content	sndppt	%	
	Silt content	sltppt	%	
	Derived available soil water capacity	wwp	–	
	Soil organic carbon content	orcdrc	$g\,kg^{-1}$	
	Soil ph	phihox	pH	Soil pH × 10 in H_2O
	Bulk density	bldfie	$kg\,m^{-3}$	
	Cation exchange capacity	cecsol	$cmolc\,kg^{-1}$	
	Coarse fragments volumetric	crfvol	%	
	Grade of a sub-soil being acid	acdwrb	pH	Grade of a sub-soil being acid e.g. having a pH < 5 and low BS
	Depth to bedrock (*r* horizon) up to 200 cm	bdricm	cm	
	Probability of occurrence of *r* horizon	bdrlog	%	
	Cumulative probability of organic soil	histpr	–	Cumulative probability of organic soil based on the TAXOUSDA and TAXNWRB
	Sodic soil grade	slgwrb	pH	Sodic soil grade based on WRB soil types and soil pH

Table 5. List of variables derived from the MERID DEM.

MERIT DEM: Multi-Error-Removed Improved-Terrain DEM v1.0.3 (Yamazaki et al., 2017)				
spatial resolution	90 m			
temporal resolution	–			
time range	–			
statistics	mean, standard deviation, minimum, maximum, range			
Variable Type	Variable	Abbreviation	Unit	Description
Elevation	elevation	elev	m	The MERIT DEM represents elevation in meters

Sect. 2.1). These tables have been integrated into a PostgreSQL database as a backbone for the GeoFRESH online platform (available at https://www.geofresh.org, last access: 4 February 2026, Domisch et al., 2024) (Fig. 1), where users can interactively retrieve the data for any location of interest. The graphical user interface allows to upload point coordinates to the portal, move (or "snap") the coordinates to the Hydrography90m stream network, and annotate the coordinates with Environment90m variables (Domisch et al., 2024).

Table 6. List of variables derived from the FLO1K streamflow dataset.

FLO1K, global maps of mean, maximum and minimum annual streamflow (Barbarossa et al., 2018)				
spatial resolution	1 km			
temporal resolution	Long term annual average			
time range	1980–2010			
statistics	mean, standard deviation, minimum, maximum, range			
Variable Type	Variable	Abbreviation	Unit	Description
Flow	streamflow	flo1k	$m^3\ s^{-1}$	The long-term mean annual flow represents the average of the year-specific FLO1K maps for mean annual flow over the period 1980–2010

Table 7. List of variables derived from the Global Aridity and Evapotranspiration dataset.

Global Aridity Index and Potential Evapotranspiration Climate Database v3 (Zomer and Trabucco, 2022)				
spatial resolution	1 km			
temporal resolution	Long Term Average			
time range	1970–2000			
statistics	mean, standard deviation, minimum, maximum, range			
Variable Type	Variable	Abbreviation	Unit	Description
	Evapotranspiration	gevapt	mm	Potential Evapotranspiration (ET0) based upon implementation of the FAO-56 Penman-Monteith Reference Evapotranspiration (ET0) equation
	Aridity index	garid	–	Ratio between precipitation and ET0. Values reported have been multiplied by a factor of 10 000.

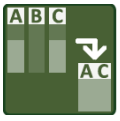
In addition, all tables were converted to follow the same tiling scheme as in the Hydrography90m dataset, such that the Environment90m and Hydrograhy90m datasets are compatible regarding the downloading and processing functionalities of the hydrographr R-package (Schürz et al., 2023). A number of new functions have been added to the hydrographr R-package specifically created to download and extract information from the original tables to specific subsets of target sub-catchments (Table 8). Although processing data frames in R is usually easy, the added value of the new functions is to overcome with challenges related to the size of the tables, especially at large geographical extents, and to process and e.g. subset the large tables efficiently using R-commands. The main feature of the hydrographr R-package is that it uses open-source third-party command-line tools without actually reading the data into R, such that large data files can be efficiently processed using R (Schürz et al., 2023). The next section presents a workflow to illustrate the use of the new available functions. Similarly, the tables in the tiling system can be manually and directly downloaded from the https://hydrography.org/environment90m (last access: 4 February 2026) website, by a point-and-click selection of the variables and tiles of interest.

3 Case study: fish habitat suitability in the Danube River Basin

The Environment90m database is especially suited for freshwater biogeographic analyses, including predictive modelling of freshwater species distributions. This task usually requires range-wide spatial and environmental data, which can become massive when working at high resolution and large spatial extents. To facilitate the acquisition and processing of the large tables of Environment90m, and to enable a seamless integration with the Hydrography90m network, we have developed additional functions in the hydrographr R-package (Schürz et al., 2023) (see also Sect. 2.3). The following case study illustrates an example workflow to create a habitat suitability map of the Danube streber *Zingel streber*, a freshwater ray-finned fish in the family Percidae (Fig. 3). We provide the workflow at Case study – Danube (Environment90m) (https://glowabio.github.io/hydrographr/articles/case_study_Danube.html, last access: 4 February 2026). The main objective of the workflow is to showcase the computational efficiency of the hydrographr R-package in processing the large environmental tables for a simplified predictive modelling example.

The first step is to identify the $20° \times 20°$ tiles that overlap with a bounding box polygon of the Danube River

Table 8. New functions added to the hydrographr R-package to download, process and extract information from the Environment90m dataset.

Category		Function name	Description
Downloading		`download_*_table()`	The various `download_*_tables()` functions allow to retrieve the Environment90m variable names and download data of the Environment90m datasets, which are split into 20° × 20° tiles.
Processing		`split_table()`	Splits a table into multiple parts of equal length by splitting it along a set number of rows.
Reading and data extraction		`get_predict_table()`	Creates a table containing the environmental variables from a specific subset of sub-catchments.
		`get_modelfit_table()`	Gets the environmental variables for each species occurrences and pseudo-absences at given point locations.

Basin by applying the function `get_tile_id()`. Since the units of analysis to model the distribution of the species are the sub-catchments, we need to download the raster files of sub-catchments for each of the tiles using the `download_tiles()` function, crop each tile to the extension of interest with the function `crop_to_extent()` and merge the pieces of each tile with the function `merge_tiles()` to obtain a final raster file of sub-catchments for the bounding box of the Danube basin.

A parallel task is to download the corresponding Environment90m tables for each tile of the selected environmental variables. There are a number of functions dedicated to download each of the available datasets (e.g. `download_landcover_tables()`). The tables will be downloaded to disk and from here, they can be subset and merged, for example to process only those sub-catchment IDs that are present in the area of interest, in our case the Danube River Basin. This processing is done with the new function `get_predict_table()` which uses as arguments (i) the path on disk where the downloaded tables are located, and (ii) the list of sub-catchment IDs which have been previously identified, using the function `extract_ids()` on the sub-catchment raster file of the area of interest. Here, either all or only a subset of the aggregation statistics (e.g., mean, range) can be selected.

The output is a large table (i.e., the so-called range-wide "prediction table") with all the sub-catchments of the area of interest and the values of the selected environmental variables. This table is still on disk, and for initial screening, a subset of it can be loaded into R to run exploratory or correlation analyses to make a selection of uncorrelated variables if the purpose is to e.g. quantify species ecological niches.

The species distribution modelling requires as input a table relating the species occurrence locations with the environmental data at those locations (i.e., the so-called "model-fit table"). The function `get_modelfit_table()` creates this table by combining (i) a table of species geographic locations (i.e., coordinates), (ii) the previously created range-wide prediction table, and (iii) the raster of sub-catchments generated during the first steps of the workflow. The model-fit table should contain the species occurrences and absences (or pseudo-absences) and their associated environmental values. The user can provide the occurrences and absences together. Alternatively, the function offers the possibility to create a user defined number of random pseudo-absences.

The model-fit table can be imported into R, where any modelling technique (e.g., Random Forest) can be applied to estimate the ecological niche of the species and predict the probability of habitat suitability of the species in the area of interest. The prediction itself is done by calling the model that has been previously fit and the "prediction table" we created before. Depending on the extent of the area of interest, the prediction table could be very large and reading the table into any statistical software or even running the prediction itself could be computationally unfeasible. To make the prediction calculations feasible, the large prediction table could then be split into several sections (using the `split_table()` function) and the prediction could be applied to each part in turn. The single output tables can be merged later. The final prediction table consists of each sub-catchment ID and its corresponding probability of occurrence value. This table can then be used with the `reclass_raster()` function to reclassify the original sub-catchment raster file to create a new raster file that shows the probability of habitat suitability of the species.

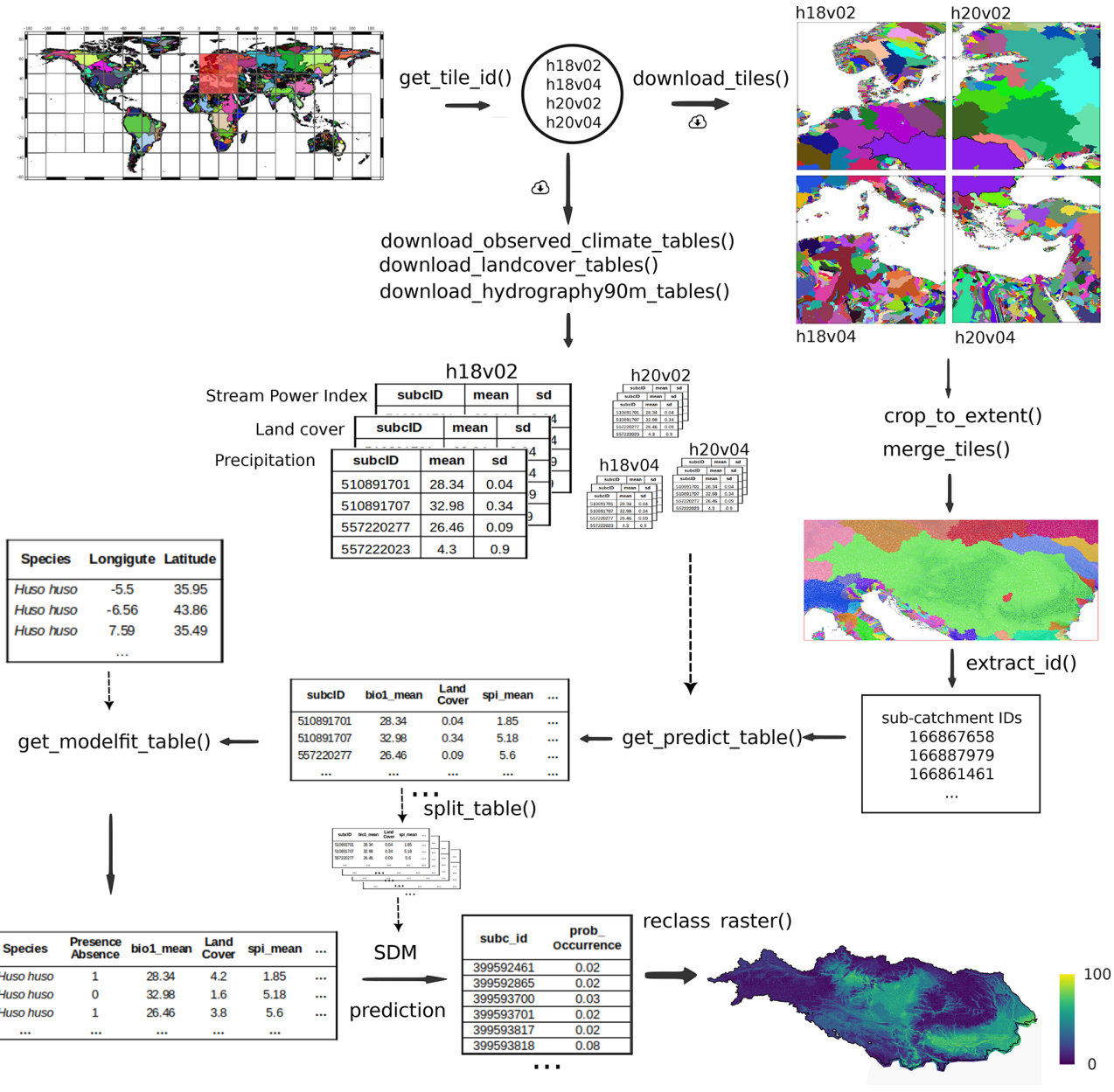


Figure 3. Workflow illustration for the case study of calculating the habitat suitability of a freshwater fish species, the Danube streber (*Zingel streber*), in the Danube River Basin. The diagram shows the new functions available in the hydrographr R-package to download the tables with the variables available in the Environment90m dataset, merge them for the area of interest and join the tables with species occurrence data as inputs to species distribution models (SDMs).

4 Case study: environmental characterization of lakes' upstream basins

We acknowledge that Environment90m focuses mainly on lotic habitats. To extend the data usage also to lentic habitats, we offer the possibility to extract Environment90m data not only for rivers but also across lakes and their contributing catchments. We provide an example code in a vignette, available at https://glowabio.github.io/hydrographr/articles/case_study_lakes.html (last access: 4 February 2026).

We have created three new functions, `extract_lake_ids()`, `get_lake_intersection()` and `get_lake_catchment()` (Table 9) that are available in the hydrographr R-package. Specifically, they allow (i) to identify the location of a lake within the Hydrography90m stream network, (ii) to obtain the intersection points

between lakes and the stream network and (iii) to delineate for each intersection point the upstream lake catchment area (Fig. 4). All functions are built on each other. For instance, the `extract_lake_ids()` function enables to retrieve lake IDs either within a specified area through a bounding box or at given point locations. The lake IDs can then be used in the `get_lake_intersection()` function specifying the lakes of interest to identify the stream cells that intersect the lakes outline and the stream network. Additionally, the resulting intersection table, containing all intersection points of the lake and stream network, includes information on flow accumulation (mean and maximum values), its coordinates and the unique stream segment ID of the Hydrography90m stream network. From this we can delineate the upstream catchment area with the `get_lake_catchment()` function for any intersection point. The intersection point with the highest flow accumulation value represents the lake outlet covering the entire upstream lake catchment area. These functions then allow to extract the environmental variables across the upstream catchment area for any lake connected to the network. For instance, by using the land cover data time series in Environment90m, it is possible to quantify the annual land cover changes in the catchment area. The vignette uses the HydroLAKES dataset (Messager et al., 2016), though the functionality is generic for any lake vector dataset. In the GeoFRESH platform, currently only HydroLAKES are available (Domisch et al., 2024).

5 Applications

The availability of globally standardized environmental data that address the structure of the hydrographic network enables comparative studies between regions. It facilitates large-scale biogeographic analyses in the freshwater realm and makes Environment90m particularly valuable for global-scale freshwater biodiversity research. For instance, a recent study focusing on the Guineo-Congolian region, a biodiversity hotspot in the Afrotropics, integrated stream network attributes derived from Environment90m with macroinvertebrate occurrence records spanning 2890 sub-catchments and stream orders 1–12, enabling biogeographic analyses in a previously understudied region. Another application of Environment90m in large-scale biodiversity assessments is demonstrated by (Haase et al., 2023), who used the dataset to analyse temporal freshwater invertebrate diversity trends across Europe. To identify environmental predictors that might drive these trends, the study used topographic, climatic and land-use variables aggregated at the sub-catchment level.

In addition, Environment90m is the backbone of a recent global study on ecological niche breadths of aquatic insect genera worldwide, where a suite of environmental predictors, such as mean stream slope gradient, stream length, bioclimatic variables and soil characteristics, was extracted per sub-catchment. These variables formed the basis for characterizing genus-level niches using the Climate-niche factor analysis (CNFA) and assessing patterns of aquatic insect niche breadth across freshwater insect assemblages (Grigoropoulou, 2024, in review). Moreover, a study from Darthmouth College focusing on how the extent of permafrost sets the drainage density in the Arctic (Vecchio et al., 2024) also used stream network variables, derived from Environment90m, to calculate the drainage density in arctic watersheds between 23.5 and 90° N latitude.

Although not using Environment90m directly, other authors have credited Hydrography90m as a source to derive and map headwaters and analyze streamflow dynamics, after using the dataset in their study on advancing science for global water protection (Golden et al., 2025). Such applications demonstrate the value of the Environment90m dataset for freshwater biodiversity research worldwide, where globally standardized data accounting for the network structure are needed.

6 Data availability

The metadata of the Environment90m dataset is stored at https://doi.org/10.18728/igb-fred-995.0 (García Márquez et al., 2025a).

The Environment90m data can be obtained from the following sources:

- The primary Environment90m data is available as zipped .csv-tables. The data comes in $20° \times 20°$ tiles, covering the same geographic extent and structure as the Hydrography90m dataset. These tiles can be interactively downloaded from https://hydrography.org/environment90m (last access: 4 February 2026).

- We recommend downloading and attaching the tables to the Hydrography90m stream network using the hydrographr R-package (Schürz et al., 2023). We provide example code at https://glowabio.github.io/hydrographr/articles/case_study_Danube.html (last access: 4 February 2026).

- For single point occurrences (i.e. coordinates), we offer the possibility to upload these to the GeoFRESH online platform (Domisch et al., 2024), available at https://geofresh.org/ (last access: 4 February 2026), and extract and download the Environment90m data either for the focal sub-catchment, or the aggregated data for the upstream contributing area.

7 Code availability

We provide all code for creating the Environment90m dataset at https://doi.org/10.5281/zenodo.18483574 (García Márquez et al., 2025b).

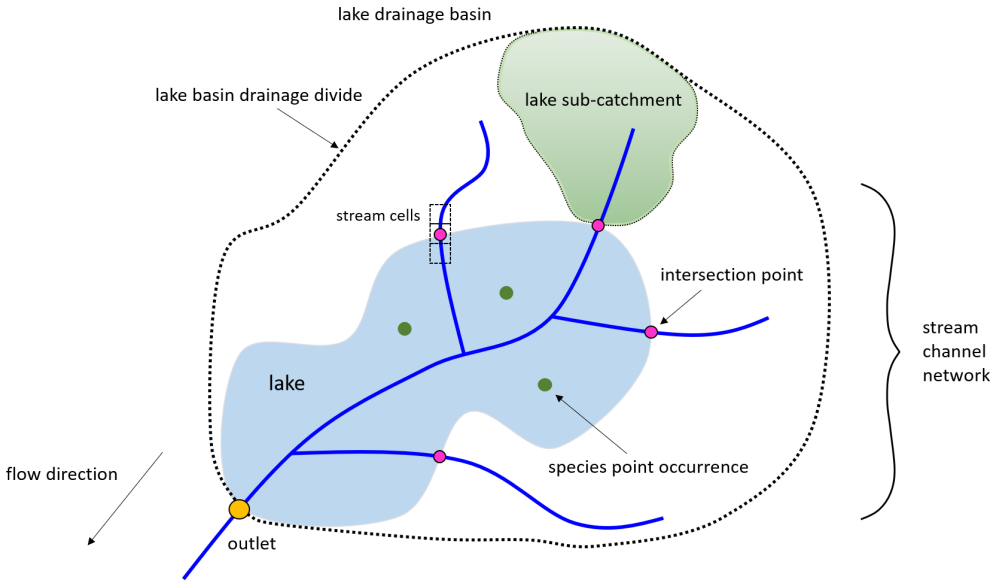


Figure 4. Schematic overview of the lake integration into the Hydrography90m stream network; accessible through the new hydrographr lake functions. See the main text for a detailed description.

Table 9. New functions added to the hydrographr R-package to integrate lakes with stream network and summarize information from the Environment90m dataset to lake upstream catchments.

Category		Function name	Description
Lake integration		`extract_lake_ids()`	Extracts the lake ID value for lakes that fall within a bounding box or that have occurrence points, using the GDAL function `gdallocationinfo`.
		`get_lake_intersection()`	Obtains the coordinates of all intersections between the lake and the stream network, along with details such as flow accumulation, using GRASS GIS and GDAL functions.
		`get_lake_catchment()`	Calculates the upstream basin for the lake outlet and/or each intersection point, using the GRASS GIS function `r.water.outlet`

8 Discussion

The Hydrography90m dataset provides globally standardized environmental information of seven different environmental datasets for a total of 104 single variables within each of the 726 million sub-catchments of the Hydrography90m dataset. The Environment90m dataset is available in different formats (see Sect. 2.3) and can be seamlessly integrated with different tools (e.g., hydrographr R-package, The GeoFRESH platform) and integrated into workflows to create and advance novel freshwater-specific studies (see vignette examples).

The use of the sub-catchment as the spatial unit of analysis to summarize the environmental information has numerous advantages for freshwater ecosystem and biodiversity analysis, in contrast to the use of e.g. raster cells, as commonly done in terrestrial biogeographic analyses. Habitats and organisms are strongly defined by the topographical and topological influences of their surrounding environment and sub-catchments encompass these fluxes. Moreover, aggregating upstream sub-catchments can provide a more comprehensive understanding of the influences that fluxes and nature-state characteristics of upstream areas have on downstream habitats. In general, the information on the connection between sub-catchments opens the door to new type of analyses where network routing, river connectivity and fragmentation processes can be considered.

We also see the calculation of summary statistics for each sub-catchment as an advantage given the increase in the va-

riety of explanatory variables that are important for understanding the distributions of freshwater habitats and biodiversity. Instead of characterizing a single sub-catchment only by the mean of a variable, e.g., elevation, the inclusion of standard deviation, or the minimum and maximum values can provide a measure of the variability of elevation in the sub-catchment. For instance, high standard deviation would represent a sub-catchment in a steep terrain, which cannot be represented only by the mean.

Although we incorporate variables that are commonly used in freshwater biogeographic analyses, there are more variables available globally that could be incorporated to the Environment90m dataset. In the near future we will add the remaining future climate projections, i.e., GCMs and time period (2011–2040). We did not include point-based data (such as nutrient or water quality measurements), and likewise focused on high-resolution environmental data. Thus, we recognize that other important variables for freshwater ecosystems may be missing in Environment90m. The challenge is, however, that coarse and gridded data, such as the global nitrogen fertilizer application map (Nishina et al., 2017) at approx. 50 km spatial resolution, can not be easily attributed to single stream segments or sub-catchments. For more regional or local studies, these type of layers would not be compatible with the Environment90m datasets. We plan to include further variables, such as the annual stream flow estimates of the FLO1K dataset (Barbarossa et al., 2018).

Taken together, we expect that Environment90m offers a unique possibility in analysing the environmental contingencies of freshwater habitats at high spatial resolution. Moreover, the dataset supports biogeographic analyses of freshwater habitats and biodiversity, and contributes towards the recent freshwater biodiversity conservation targets by providing a solid and globally standardized baseline of high-resolution environmental information.

Author contributions. JGM and SD designed the study. JGM developed and implemented the workflow and processing chain in the Yale-HPC to compute the Environment90m data. MB processed the data for the download and added the download functionality to the https://hydrography.org (last access: 4 February 2026). VB added the Environment90m data to the PostgreSQL database. YTC, VB and AG added the download functionality to the GeoFRESH online platform. JGM, MB, AG, MS, TT and YTC wrote the functions to download and process the data in the hydrographr R-package. All authors discussed the results, and all authors contributed to the writing of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. We thank the Yale Center for Research Computing for guidance and use of the research computing infrastructure. We also thank the reviewers for their constructive comments which helped us to improve the manuscript.

Financial support. This research has been supported by the Leibniz Competition project "Global freshwater biodiversity, biogeography and conservation" (https://glowabio.org/, last access: 4 February 2026; grant no. J45/2018). This work received also funding by the German Research Foundation (NFDI4Earth, DFG340 project no. 460036893, https://www.nfdi4earth.de, last access: 4 February 2026; and NFDI4Biodiversity, DFG project no. 442032008, https://www.nfdi4biodiversity.org, last access: 4 February 2026), and DFG reference number DO 1880/6-1, project no. 533943718. We also received funding by the European Commission's Horizon Europe Research and Innovation programme under grant nos. 101094434 (AquaINFRA), 101059264 (SOS-Water – Towards defining a safe operating space for the entire water resources in a changing climate and society), 101093985 (DANUBE4all). Yusdiel Torres-Cambas received funding by the Alexander von Humboldt Foundation (Ref. 3.2-CUB-1212347-GF-P) and DFG (CRC RESIST, SFB 1439/1345 2021–426547801). Vanessa Bremerich received funding through the Leibniz Competition project "Freshwater Megafauna Futures" and DFG (CRC RESIST, SFB 1439/1 2021–426547801). Kristi Bego was funded by the German Academic Exchange Service (DAAD reference no. 91902112).

Review statement. This paper was edited by Christine I. B. Wallis and reviewed by Anne Lewerentz and one anonymous referee.

References

Almeida, B. and Cabral, P.: Data-Driven Modelling of Freshwater Ecosystems: A Multiscale Framework Based on Global Geospatial Data, in: GISTAM, 104–111, ISBN 978-989-758-649-1, 2023.

Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M. M., Shen, L. Q., and Domisch, S.: Hydrography90m: a new high-resolution global hydrographic dataset, Earth System Science Data, 14, 4525–4550, https://doi.org/10.5194/essd-14-4525-2022, 2022.

Arrouays, D., Lelyk, G., Hempel, J. W., McKenzie, N. J., Lagacherie, P., Lelyk, G., McBratney, A. B., Odeh, I. O. A., Sanchez, P. A., Thompson, J. A., Grundy, M. G., Hartemink, A. E., Hong, S. Y., and Zhang, G.-L.: GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, in: Advances in Agronomy, vol. 125, Academic Press, 93–134, https://doi.org/10.1016/B978-0-12-800137-0.00003-0, 2014.

Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H., and Schipper, A. M.: FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015, Scientific Data, 5, 180052, https://doi.org/10.1038/sdata.2018.52, 2018.

Basen, T., Chucholl, C., Oexle, S., Ros, A., and Brinker, A.: Suitability of Natura 2000 sites for threatened freshwater species under projected climate change, Aquatic Conservation: Marine and Freshwater Ecosystems, 32, 1872–1887, https://doi.org/10.1002/aqc.3899, 2022.

Bellin, N., Tesi, G., Marchesani, N., and Rossi, V.: Species distribution modeling and machine learning in assessing the potential distribution of freshwater zooplankton in Northern Italy, Ecological Informatics, 69, 101682, https://doi.org/10.1016/j.ecoinf.2022.101682, 2022.

Benda, L., Poff, N. L., Miller, D., Dunne, T., Reeves, G., Pess, G., and Pollock, M.: The network dynamics hypothesis: how channel networks structure riverine habitats, BioScience, 54, 413–427, 2004.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Brunner, A., Márquez, J. R. G., and Domisch, S.: Downscaling future land cover scenarios for freshwater fish distribution models under climate change, Limnologica, 104, 126139, https://doi.org/10.1016/j.limno.2023.126139, 2024.

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

Domisch, S., Bremerich, V., Buurman, M., Kaminke, B., Tomiczek, T., Torres-Cambas, Y., Grigoropoulou, A., Garcia Marquez, J., Amatulli, G., Grossart, H.-P., Gessner, M., Mehner, T., Adrian, R., and De Meester, L.: GeoFRESH – an online platform for freshwater geospatial data processing, International Journal of Digital Earth, 17, 2391033, https://doi.org/10.1080/17538947.2024.2391033, 2024.

Ebi, K. L., Hallegatte, S., Kram, T., Arnell, N. W., Carter, T. R., Edmonds, J., Kriegler, E., Mathur, R., O'Neill, B. C., Riahi, K., Harald Winkler, H., Van Vuuren, D., and Zwickel, T.: A new scenario framework for climate change research: background, process, and future directions, Climatic Change, 122, 363–372, 2014.

ESA: Land Cover CCI Product User Guide Version 2. Tech. Rep., Tech. rep., European Space Agency, https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (last access: 4 February 2026), 2017.

Friedrichs-Manthey, M., Langhans, S. D., Hein, T., Borgwardt, F., Kling, H., Jähnig, S. C., and Domisch, S.: From topography to hydrology – The modifiable area unit problem impacts freshwater species distribution models, Ecology and Evolution, 10, 2956–2968, 2020.

García Márquez, J. R., Grigoropoulou, A., Tomiczek, T., Schürz, M., Bremerich, V., Torres-Cambas, Y., Buurman, M., Bego, K., Amatulli, G., and Domisch, S.: Environment90m – globally standardized environmental variables for freshwater science at high spatial resolution, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) [data set], https://doi.org/10.18728/igb-fred-995.0, 2025a.

García Márquez, J. R., Grigoropoulou, A., Tomiczek, T., Schürz, M., Bremerich, V., Torres-Cambas, Y., Buurman, M., Bego, K., Amatulli, G., Domisch, S.: Environment90m – globally standardized environmental variables for freshwater science at high spatial resolution, Zenodo [code], https://doi.org/10.5281/zenodo.18483574, 2025b.

Golden, H. E., Christensen, J. R., McMillan, H. K., Kelleher, C. A., Lane, C. R., Husic, A., Li, L., Ward, A. S., Hammond, J., Seybold, E. C., Jaeger, K. L., Zimmer, M., Sando, R., Jones, C. N., Segura, C., Mahoney, D. T., Price, A. N., and Cheng, F.: Advancing the science of headwater streamflow for global water protection, Nature Water, 3, 16–26, https://doi.org/10.1038/s44221-024-00351-1, 2025.

GRASS Development Team: Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.4, Open Source Geospatial Foundation, USA, Zenodo [code], https://doi.org/10.5281/zenodo.5176030, 2024.

Grigoropoulou, A.: Open science approaches on assessing global aquatic insect biodiversity, Dissertation, Freie Universität Berlin, https://doi.org/10.17169/refubium-45697, 2024.

Grigoropoulou, A., Hamid, S. A., Acosta, R., Akindele, E. O., Al-Shami, S. A., Altermatt, F., Amatulli, G., Angeler, D. G., Arimoro, F. O., Aroviita, J., Astorga-Roine, A., Bastos, R. C., Bonada, N., Boukas, N., Brand, C., Bremerich, V., Bush, A., Cai, Q., Callisto, M., Chen, K., Cruz, P. V., Dangles, O., Death, R., Deng, X., Domínguez, E., Dudgeon, D., Eriksen, T. E., Faria, A. P. J., Feio, M. J., Fernández-Aláez, C., Floury, M., García-Criado, F., García-Girón, J., Graf, W., Grönroos, M., Haase, P., Hamada, N., He, F., Heino, J., Holzenthal, R., Huttunen, K.-L., Jacobsen, D., Jähnig, S. C., Jetz, W., Johnson, R. K., Juen, L., Kalkman, V., Kati, V., Keke, U. N., Koroiva, R., Kuemmerlen, M., Langhans, S. D., Ligeiro, R., Van Looy, K., Maasri, A., Marchant, R., Garcia Marquez, J. R., Martins, R. T., Melo, A. S., Metzeling, L., Miserendino, M. L., Moe, S. J., Molineri, C., Muotka, T., Mustonen, K.-R., Mykrä, H., Cavalcante do Nascimento, J. M., Valente-Neto, F., Neu, P. J., Nieto, C., Pauls, S. U., Paulson, D. R., Rios-Touma, B., Rodrigues, M. E., de Oliveira Roque, F., Salazar Salina, J. C., Schmera, D., Schmidt-Kloiber, A., Shah, D. N., Simaika, J. P., Siqueira, T., Tachamo-Shah, R. D., Theischinger, G., Thompson, R., Tonkin, J. D., Torres-Cambas, Y., Townsend, C., Turak, E., Twardochleb, L., Wang, B., Yanygina, L., Zamora-Muñoz, C., and Domisch, S.: The global EPTO database: Worldwide occurrences of aquatic insects, Global Ecology and Biogeography, 32, 642–655, https://doi.org/10.1111/geb.13648, 2023.

Guareschi, S., Cancellario, T., Oficialdegui, F. J., and Clavero, M.: Insights from the past: Invasion trajectory and niche trends of a global freshwater invader, Global Change Biology, 30, e17059, https://doi.org/10.1111/gcb.17059, 2024.

Haase, P., Bowler, D. E., Baker, N. J., Bonada, N., Domisch, S., Garcia Marquez, J. R., Heino, J., Hering, D., Jähnig, S. C., Schmidt-Kloiber, A., Stubbington, R., Altermatt, F., Álvarez Cabria, M., Amatulli, G., Angeler, D. G., Archambaud-Suard, G., Jorrín, I. A., Aspin, T., Azpiroz, I., Bañares, I., Ortiz, J. B., Bodin, C. L., Bonacina, L., Bottarin, R., Cañedo-Argüelles, M., Csabai, Z., Datry, T., de Eyto, E., Dohet, A., Dörflinger, G., Drohan, E., Eikland, K. A., England, J., Eriksen, T. E., Evtimova, V.,

Feio, M. J., Ferréol, M., Floury, M., Forcellini, M., Forio, M. A. E., Fornaroli, R., Friberg, N., Fruget, J.-F., Georgieva, G., Goethals, P., Graça, M. A. S., Graf, W., House, A., Huttunen, K.-L., Jensen, T. C., Johnson, R. K., Jones, J. I., Kiesel, J., Kuglerová, L., Larrañaga, A., Leitner, P., L'Hoste, L., Lizée, M.-H., Lorenz, A. W., Maire, A., Arnaiz, J. A. M., McKie, B. G., Millán, A., Monteith, D., Muotka, T., Murphy, J. F., Ozolins, D., Paavola, R., Paril, P., Peñas, F. J., Pilotto, F., Polášek, M., Rasmussen, J. J., Rubio, M., Sánchez-Fernández, D., Sandin, L., Schäfer, R. B., Scotti, A., Shen, L. Q., Skuja, A., Stoll, S., Straka, M., Timm, H., Tyufekchieva, V. G., Tziortzis, I., Uzunov, Y., van der Lee, G. H., Vannevel, R., Varadinova, E., Várbíró, G., Velle, G., Verdonschot, P. F. M., Verdonschot, R. C. M., Vidinova, Y., Wiberg-Larsen, P., and Welti, E. A. R.: The recovery of European freshwater biodiversity has come to a halt, Nature, 620, 582–588, 2023.

He, C., Yang, C.-J., Turowski, J. M., Ott, R. F., Braun, J., Tang, H., Ghantous, S., Yuan, X., and Stucky de Quay, G.: A global dataset of the shape of drainage systems, Earth System Science Data, 16, 1151–1166, https://doi.org/10.5194/essd-16-1151-2024, 2024.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hermoso, V.: Restoring free-flowing rivers: Planning for longitudinal and lateral connectivity recovery, Journal of Applied Ecology, https://doi.org/10.1111/1365-2664.70095, 2025.

Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., and Thornbrugh, D. J.: The Stream-Catchment (StreamCat) Dataset: A database of watershed metrics for the conterminous United States, JAWRA Journal of the American Water Resources Association, 52, 120–128, 2016.

Hill, R. A., Weber, M. H., Debbout, R. M., Leibowitz, S. G., and Olsen, A. R.: The Lake-Catchment (LakeCat) Dataset: characterizing landscape features for lake basins within the conterminous USA, Freshwater Science, 37, 208–221, 2018.

Hughes, A. C.: The Post-2020 Global Biodiversity Framework: How did we get here, and where do we go next?, Integrative Conservation, 2, 1–9, https://doi.org/10.1002/INC3.16, 2023.

Jelinski, D. E. and Wu, J.: The modifiable areal unit problem and implications for landscape ecology, Landscape Ecology, 11, 129–140, 1996.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, Scientific Data, 4, 1–20, 2017.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, EnviDat [data set], https://doi.org/10.16904/envidat.228, 2021.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, Transactions American Geophysical Union, 89, 93–94, 2008.

Lehner, B., Messager, M. L., Korver, M. C., and Linke, S.: Global hydro-environmental lake characteristics at high spatial resolution, Scientific Data, 9, 351, https://doi.org/10.1038/s41597-022-01425-z, 2022.

Leopold, L. B., Wolman, M. G., Miller, J. P., and Wohl, E. E.: Fluvial Processes in Geomorphology, Courier Dover Publications, ISBN 978-0-486-84552-4, 2020.

Linke, S., Pressey, R. L., Bailey, R. C., and Norris, R. H.: Management options for river conservation planning: condition and conservation re-visited, Freshwater Biology, 52, 918–938, 2007.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Scientific Data, 6, 283, https://doi.org/10.1038/s41597-019-0300-6, 2019.

Liu, X., Yu, L., Si, Y., Zhang, C., Lu, H., Yu, C., and Gong, P.: Identifying patterns and hotspots of global land cover transitions using the ESA CCI Land Cover dataset, Remote Sensing Letters, 9, 972–981, https://doi.org/10.1080/2150704X.2018.1500070, 2018.

Lowe, W. H. and Likens, G. E.: Moving Headwater Streams to the Head of the Class, BioScience, 55, 196–197, https://doi.org/10.1641/0006-3568(2005)055[0196:MHSTTH]2.0.CO;2, 2005.

Lv, Y., Jia, L., Menenti, M., Zheng, C., Lu, J., Jiang, M., Chen, Q., and Zhang, Y.: Estimations of Dynamic Water Depth and Volume of Global Lakes Using Machine Learning, Remote Sensing, 17, https://doi.org/10.3390/rs17061052, 2025.

Maasri, A., Thorp, J. H., Gelhaus, J. K., Tromboni, F., Chandra, S., and Kenner, S. J.: Communities associated with the Functional Process Zone scale: A case study of stream macroinvertebrates in endorheic drainages, Science of the Total Environment, 677, 184–193, 2019.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A.: NHDPlus Version 2: User Guide, Tech. rep., US Geological Survey, https://www.epa.gov/system/files/documents/2023-04/NHDPlusV2_User_Guide.pdf (last access: 16 June 2025), 2012.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nature Communications, 7, 13603, https://doi.org/10.1038/ncomms13603, 2016.

Modi, P., Revel, M., and Yamazaki, D.: Multivariable Integrated Evaluation of Hydrodynamic Modeling: A Comparison of Performance Considering Different Baseline Topography Data, Water Resources Research, 58, e2021WR031819, https://doi.org/10.1029/2021WR031819, 2022.

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., and Ghazali, A. H. B.: Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS, Geomatics, Natural Hazards and Risk, 8, 1080–1102, https://doi.org/10.1080/19475705.2017.1294113, 2017.

Moore, I. D., Grayson, R. B., and Ladson, A. R.: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications, Hydrological Processes, 5, 3–30, https://doi.org/10.1002/hyp.3360050103, 1991.

Neteler, M., Bowman, M. H., Landa, M., and Metz, M.: GRASS GIS: A multi-purpose open source GIS, Environmental Modelling & Software, 31, 124–130, 2012.

Nishina, K., Ito, A., Hanasaki, N., and Hayashi, S.: Reconstruction of spatially detailed global map of NH_4^+ and NO_3^- application in synthetic nitrogen fertilizer, Earth System Science Data, 9, 149–162, https://doi.org/10.5194/essd-9-149-2017, 2017.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W.: The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, Global Environmental Change, 42, 169–180, 2017.

Schürz, M., Grigoropoulou, A., García Márquez, J., Torres-Cambas, Y., Tomiczek, T., Floury, M., Bremerich, V., Schürz, C., Amatulli, G., Grossart, H.-P., and Domisch, S.: hydrographr: An R package for scalable hydrographic data processing, Methods in Ecology and Evolution, 14, 2953–2963, 2023.

Schürz, M., Márquez, J. G., and Domisch, S.: The Scale-Dependency in Freshwater Habitat Regionalisation Analyses, Ecohydrology, 18, e70018, https://doi.org/10.1002/eco.70018, 2025.

Stoffers, T., Altermatt, F., Baldan, D., Bilous, O., Borgwardt, F., Buijse, A. D., Bondar-Kunze, E., Cid, N., Erős, T., Ferreira, M. T., Funk, A., Haidvogl, G., Hohensinner, S., Kowal, J., Nagelkerke, L. A. J., Neuburg, J., Peller, T., Schmutz, S., Singer, G. A., Unfer, G., Vitecek, S., Jähnig, S. C., and Hein, T.: Reviving Europe's rivers: Seven challenges in the implementation of the Nature Restoration Law to restore free-flowing rivers, WIREs Water, 11, e1717, https://doi.org/10.1002/wat2.1717, 2024.

Strahler, A. N.: Quantitative analysis of watershed geomorphology, Eos, Transactions American Geophysical Union, 38, 913–920, 1957.

Tan, M. L., Tew, Y. L., Chun, K. P., Samat, N., Shaharudin, S. M., Mahamud, M. A., and Tangang, F. T.: Improvement of the ESA CCI Land cover maps for water balance analysis in tropical regions: A case study in the Muda River Basin, Malaysia, Journal of Hydrology: Regional Studies, 36, 100837, https://doi.org/10.1016/j.ejrh.2021.100837, 2021.

Thorp, J. H.: Metamorphosis in river ecology: from reaches to macrosystems, Freshwater Biology, 59, 200–210, 2014.

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., Cooke, S. J., Dalton, J., Darwall, W., Edwards, G., Harrison, I., Hughes, K., Jones, T., Leclère, D., Lynch, A. J., Leonard, P., McClain, M. E., Muruven, D., Olden, J. D., Ormerod, S. J., Robinson, J., Tharme, R. E., Thieme, M., Tockner, K., Wright, M., and Young, L.: Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan, BioScience, 70, 330–342, https://doi.org/10.1093/biosci/biaa002, 2020.

Tomiczek, T., Bremerich, V., Grigoropoulou, A., Schürz, M., García Márquez, J. R., Torres-Cambas, Y., Amatulli, G., Ogashawara, I., Grossart, H.-P., Friedrichs-Manthey, M., and Domisch, S.: Towards a seamless geospatial representation of freshwater habitats, Zenodo, https://doi.org/10.5281/zenodo.14361453, 2024.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., and Cushing, C. E.: The River Continuum Concept, Canadian Journal of Fisheries and Aquatic Sciences, 37, 130–137, https://doi.org/10.1139/f80-017, 1980.

Vecchio, J. D., Palucis, M. C., and Meyer, C. R.: Permafrost extent sets drainage density in the Arctic, Proceedings of the National Academy of Sciences, 121, e2307072120, https://doi.org/10.1073/pnas.2307072120, 2024.

Ward, J. V. and Stanford, J. A.: The serial discontinuity concept of lotic ecosystems, in: Dynamics of Lotic Ecosystems, edited by: Fontaine, T. and Bartell, S. M., Ann Arbor Science Publishers, 29–42, https://www.researchgate.net/publication/247493939_The_Serial_Discontinuity_Concept_Of_Lotic_Ecosystems (last access: 4 February 2026), 1983.

Ward, J. and Stanford, J.: The serial discontinuity concept: extending the model to floodplain rivers, Regulated Rivers: Research & Management, 10, 159–168, 1995.

WWF: Living Planet Report 2020. Bending the curve of biodiversity loss: a deep dive into freshwater, WWF, https://wwfint.awsassets.panda.org/downloads/lpr_2020___deep_dive_into_freshwater___spreads_embargo_10_09_20_1_1.pdf (last access: 4 February 2026), 2020.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations: Accurate Global Terrain Elevation map, Geophysical Research Letters, 44, 5844–5853, https://doi.org/10.1002/2017GL072874, 2017.

Zomer, R. J. and Trabucco, A.: Version 3 of the "Global Aridity Index and Potential Evapotranspiration (ET0) Database": Estimation of Penman-Monteith Reference Evapotranspiration, CGIAR-CSI GeoPortal, Tech. rep., CGIAR-CSI, https://cgiarcsi.community/2019/01/24/global-aridity-index-and-potential-evapotranspiration-climate-database-v3 (last access: 4 February 2026), 2022.