

Hydrography90m - Distance to outlet

# Tools to address the longitudinal connectivity in spatial freshwater biodiversity science

S. Domisch, M. Schürz, A. Grigoropoulou, V. Bremerich, C. Schürz, G. Amatulli, Y. Torres Cambas, T. Tomiczek, M. Buurman, M. Konkol, H.-P. Grossart, J. Garcia Marquez

Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)

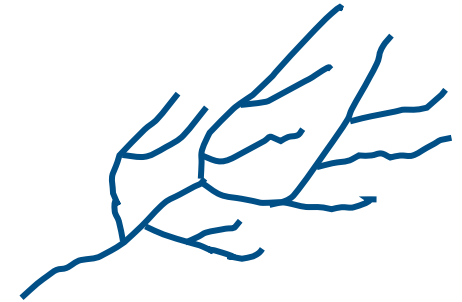
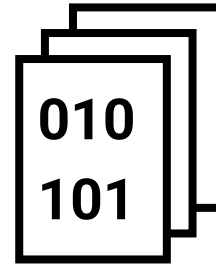
# Addressing the connectivity in large-scale freshwater geospatial workflows

- Complex geo-spatial data processing
- Integration of connectivity and routing
- Large amount of data
- Scalability across large spatio-temporal scales

# How to address these limitations?

Support scientific communities with

- **scalable processing workflows**

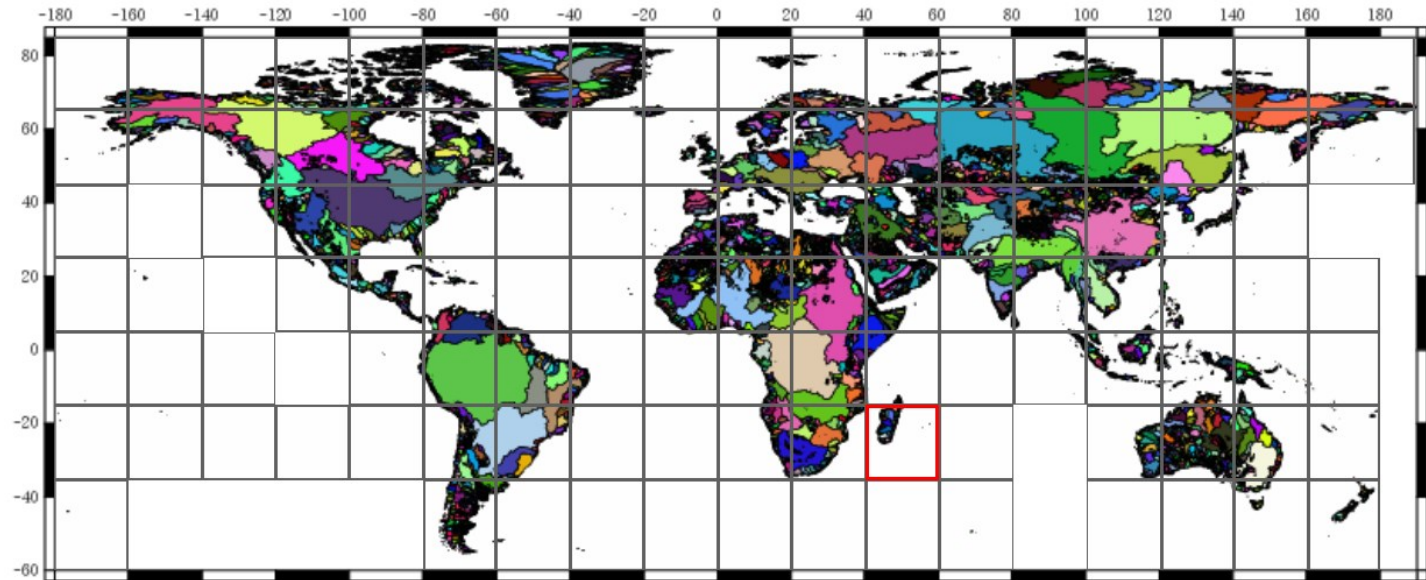


- **easy-to-use-tools for geospatial data processing**



# The Hydrography90m dataset

- 1.6 million drainage basins
- 726 million unique stream segments and sub-catchments

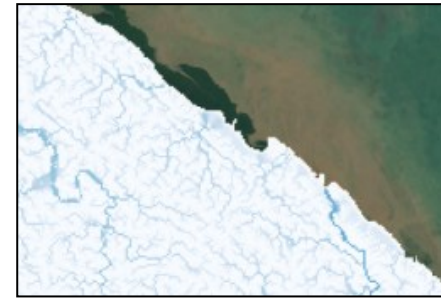


<https://hydrography.org/>

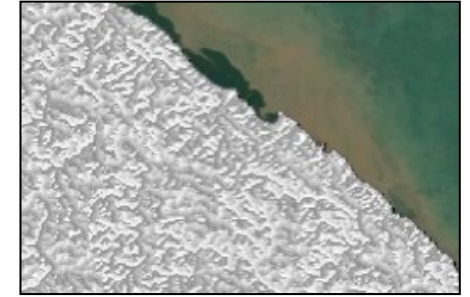


# The Hydrography90m dataset

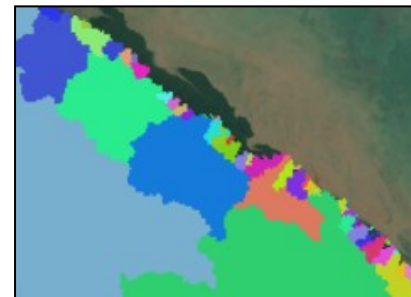
- 1.6 million drainage basins
- 726 million unique stream segments and sub-catchments
- 42 stream topographic variables
- Only open-source tools!



Accumulation



Direction



Basins



Network



Sub-catchments



Yale Center for Research Computing

<https://hydrography.org/>

# Open-source toolbox

Scalable, command-line tools



Scientific communities

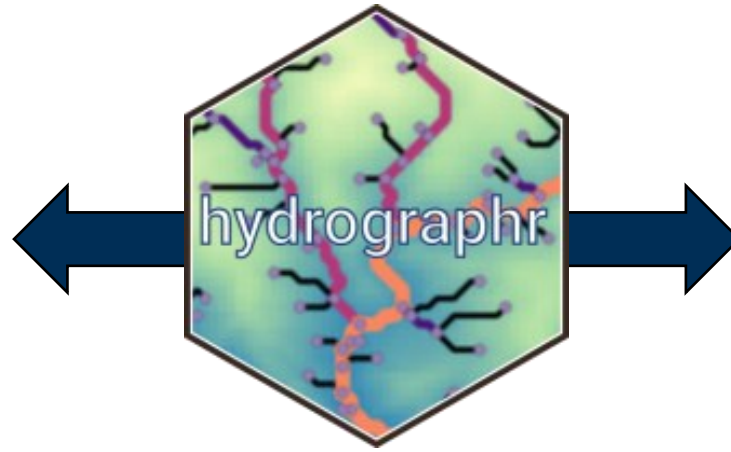


# Open-source toolbox

Scalable, command-line tools



Scientific communities



- Wrapper functions
- Scalability
- R language

# Features of the hydrographr package



- Supports freshwater geospatial data processing
- Raster, vector, graph, data.table
- Hydrography90m dataset ([hydrography.org](https://hydrography.org))
- Data processing outside R --> allows scalable analyses, handles massive data
- Currently 24 functions for
  - downloading
  - data processing
  - reading & extracting data
  - distance & connectivity
  - network structure
- Parallelization within / outside R





# hydrographr functions – Downloading data

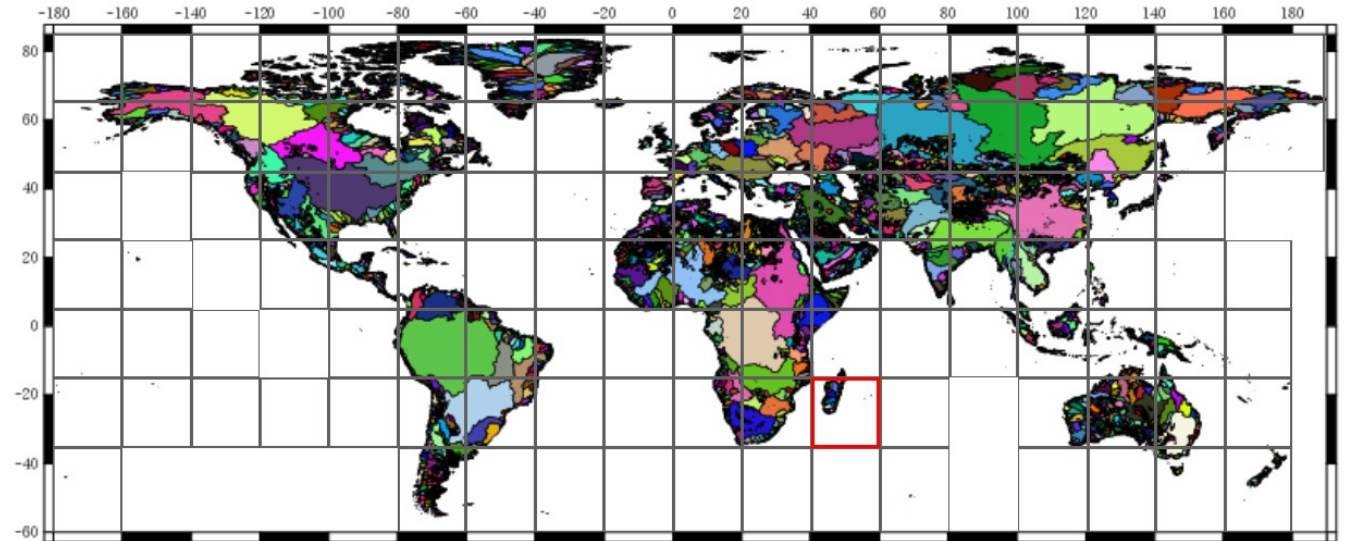


- Helper functions for downloading Hydrography90m data
- Start with e.g. species point data to automatically derive the corresponding network data

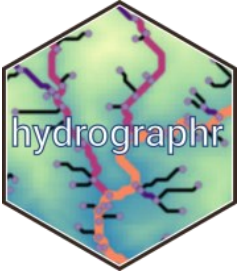
`download_tiles()`



`download_test_data()`

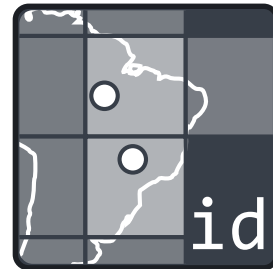


# hydrographr functions – Downloading data

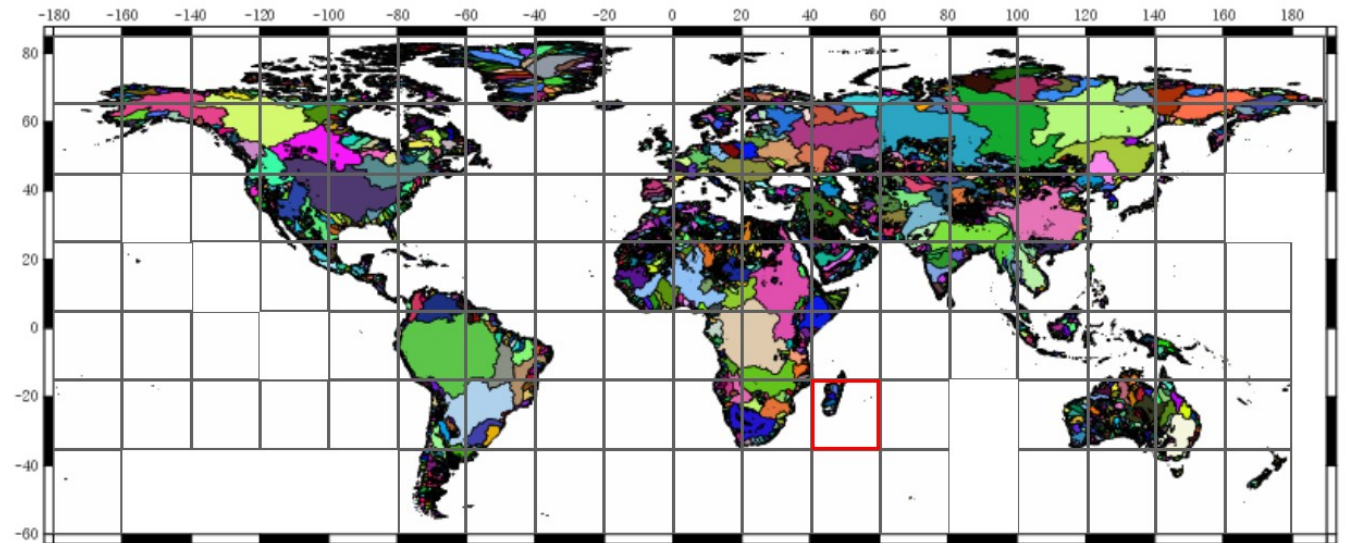


- Helper functions for downloading Hydrography90m data
- Start with e.g. species point data to automatically derive the corresponding network data

`get_tile_id()`



`get_regional_unit_id()`

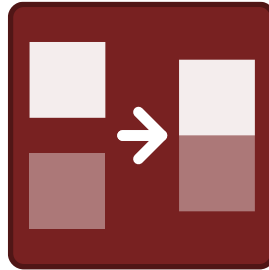


# hydrographr functions – Data processing

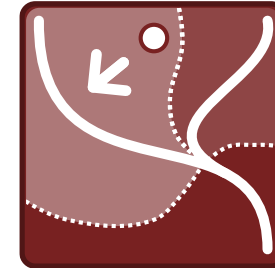


Heavy data processing is done outside R, allowing to address large study areas

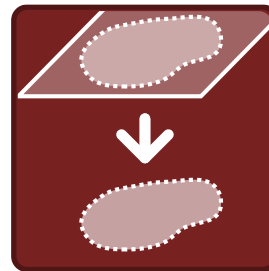
`merge_tiles()`



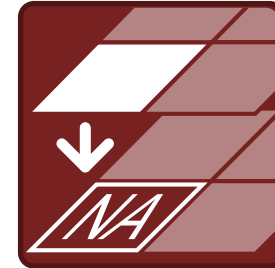
`snap_to_subc_segment()`



`crop_to_extent()`



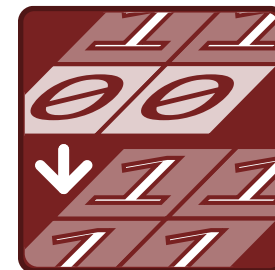
`set_no_data()`



`snap_to_network()`



`reclass_raster()`

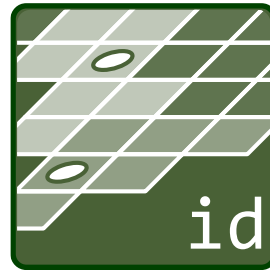


# hydrographr functions – Reading & extracting data

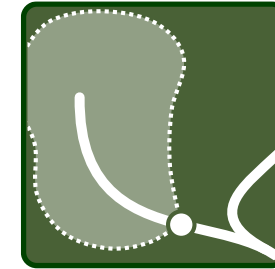


Reading & extracting large amounts of data

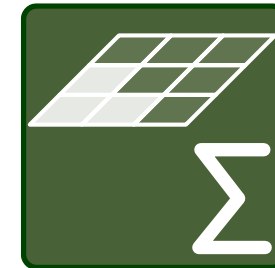
`extract_ids()`



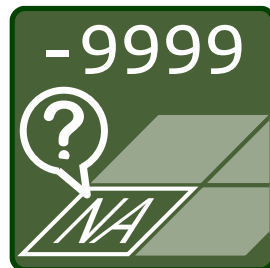
`get_upstream_catchment()`



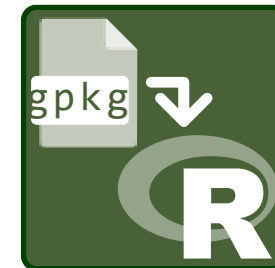
`extract_zonal_stat()`



`report_no_data()`



`read_geopackage()`

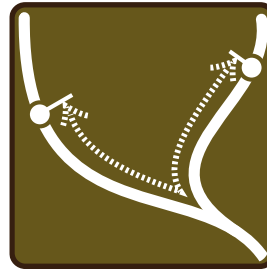


# hydrographr functions – Distance & connectivity

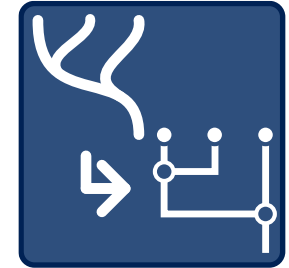


Assess point distances along the network & use graphs to analyse the network

`get_distance()`



`get_catchment_graph()`



`get_segment_neighbours()`



`get_pfafstetter_basins()`



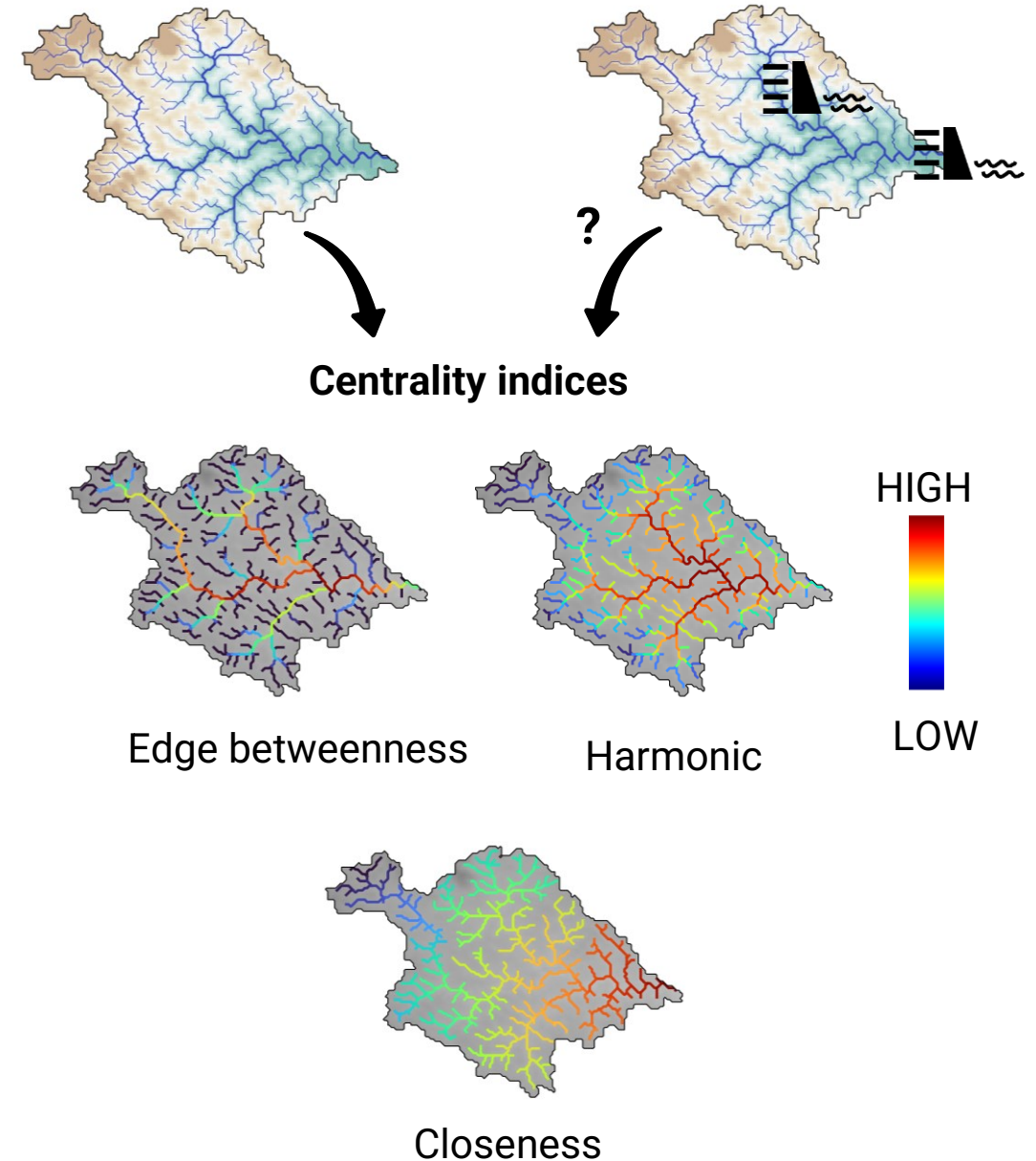
`get_centrality()`



# Connectivity & network structure

Across large spatial scales, assess the

- impact of dams on network structure
- longest-possible network distances (“free-flowing rivers”)
- degree of fragmentation
- estimated impact of new dams, or dam removal?
- effects on species populations & distributions
- ...



# hydrographr



DOI [10.5281/zenodo.7998072](https://doi.org/10.5281/zenodo.7998072) repo status **Active**

`hydrographr` provides a collection of `R` function wrappers for GDAL and GRASS-GIS functions to efficiently work with `Hydrography90m` and spatial biodiversity data. The easy-to-use functions process large raster and vector data directly on disk in parallel, such that the memory of `R` does not get overloaded. This allows creating scalable data processing and analysis workflows in `R`, even though the data is not processed directly in `R`.

We invite users to test the package and to provide feedback. Please notify us of any possible issues, bugs and feature requests under the [issues tab](#) on the top of this page.

## Installation

Please see the installation guide of the required tools at <https://glowabio.github.io/hydrographr/articles/hydrographr.html>. Afterwards, use the following lines to install the package in `R`:

```
install.packages("remotes")
remotes::install_github("glowabio/hydrographr")
library(hydrographr)
```

The pdf manual of the `hydrographr` package can be downloaded [here](#).

We thank [NFDI4Biodiversity](#) and [NFDI4Earth](#) for providing the funding that helped us getting the hydrographr package together!



# hydrographr



DOI [10.5281/zenodo.7998072](https://doi.org/10.5281/zenodo.7998072) repo status **Active**

`hydrographr` provides a collection of `R` function wrappers for GDAL and GRASS-GIS functions to efficiently work with `Hydrography90m` and spatial biodiversity data. The easy-to-use functions process large raster and vector data directly on disk in parallel, such that the memory of R does not get overloaded. This allows creating scalable data processing and analysis workflows in R, even though the data is not processed directly in R.

We invite users to test the package and to provide feedback. Please notify us of any possible issues, bugs and feature requests under the [issues tab](#) on the top of this page.

## Installation

Please see the installation guide of the required tools at <https://glowabio.github.io/hydrographr/articles/hydrographr.html>. Afterwards, use the following lines to install the package in R:

```
install.packages("remotes")
remotes::install_github("glowabio/hydrographr")
library(hydrographr)
```

The pdf manual of the `hydrographr` package can be downloaded [here](#).

We thank [NFDI4Biodiversity](#) and [NFDI4Earth](#) for providing the funding that helped us getting the hydrographr package together!



# One-time installation of the software

Setting up the package requirements | glowabio.github.io/hydrographr/articles/windows\_system\_setup.html

hydrographr 1.0.14 | Get started | Reference | Articles

Search for

## Setting up the package requirements on Windows

Source: [vignettes/windows\\_system\\_setup.Rmd](#)

To handle large geodata efficiently the `hydrographr` package relies on a suite of Open Source geoprocessing tools like `GRASS GIS`, `GDAL`, and `pktools`. Since the `hydrographr` package is optimized to run in a Linux operating system, in a Windows environment the inbuilt Windows Subsystem for Linux (WSL) is used to install a Ubuntu operating system with the additional geoprocessing packages.

Below you will find a short summary of the necessary steps to install WSL, Ubuntu and the required GIS tools.

To learn more about the WSL check out the [Microsoft documentation](#), which provides background information and step to step tutorials.

### Prerequisites

- Make sure you have at least 4 GB of free disk space.
- You need to have administrator privileges for Windows.
- You must be running Windows 10 version 2004 and higher or Windows 11, otherwise WSL is not available.

### Installation of the Windows Subsystem for Linux with Ubuntu

Open the Windows Command Prompt (`cmd.exe`) in administrator mode by clicking the Start button and typing "`cmd`" into the search bar. Then right-click "Command Prompt" and select "Run as Administrator" or highlight the result with the arrow keys and press `Ctrl+Shift+Enter` to open the command prompt with administrative privileges.

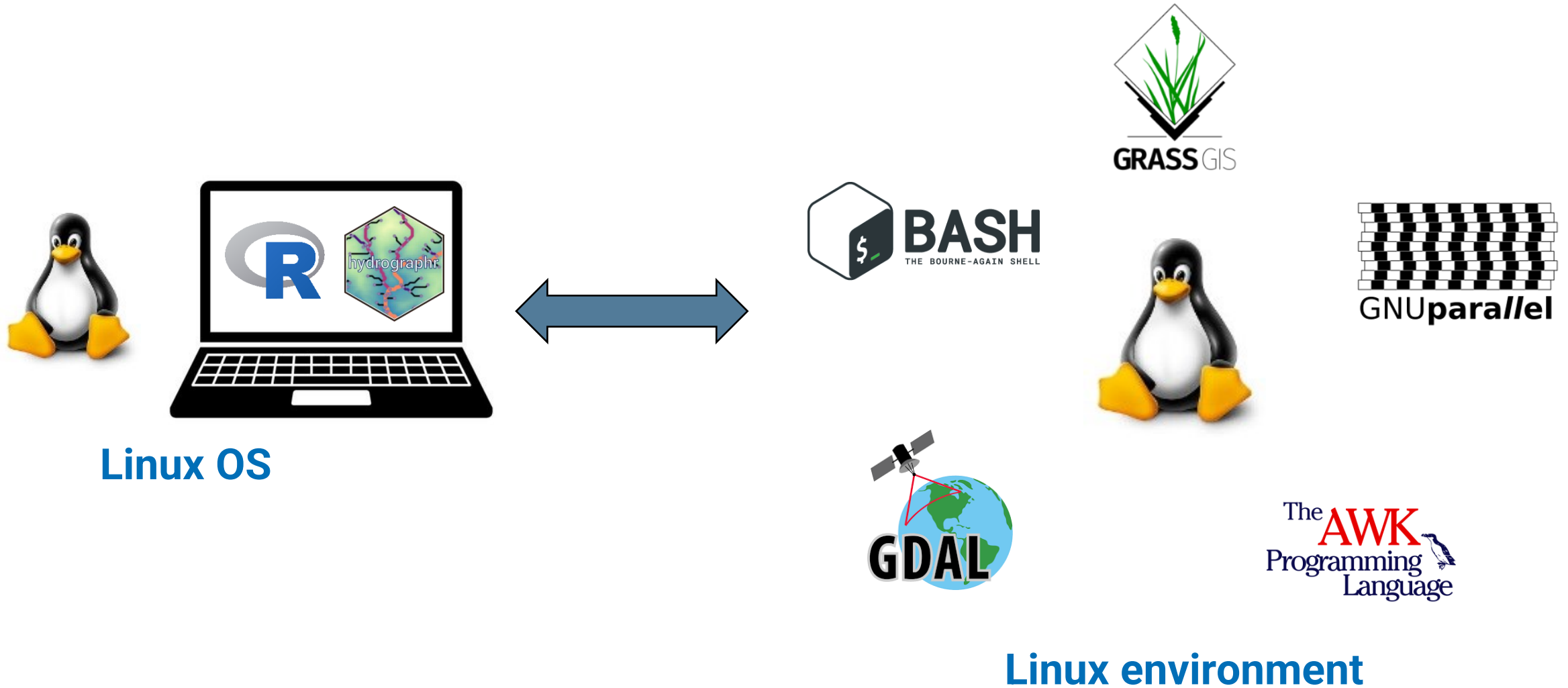
Enter `wsl --install` to enable the features necessary to run WSL and also directly install the Ubuntu distribution of Linux, which is the default option in WSL.

### On this page

- Prerequisites
- Installation of the Windows Subsystem for Linux with Ubuntu
- Installation of the required GIS tools on the WSL

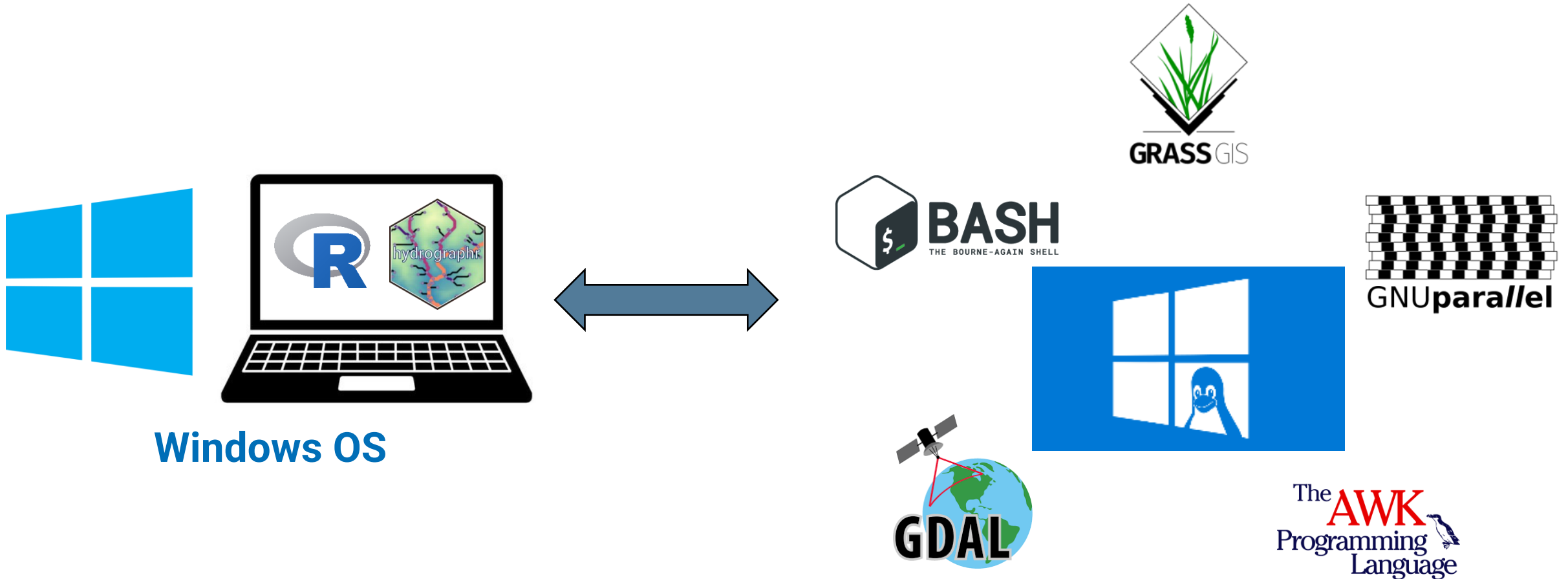
# glowabio.github.io/hydrographr/

# Data processing with open-source tools





# Data processing with open-source tools



Windows OS

Windows Subsystem for Linux (WSL)



## Bookmarks x

- crop\_to\_extent
- download\_test\_data
- download\_tiles
- extract\_from\_gpkg
- extract\_ids
- extract\_zonal\_stat
- get\_catchment\_graph
- get\_distance
- get\_pfafstetter\_basins
- get\_regional\_unit\_id
- get\_segment\_neighbours
- get\_tile\_id
- get\_upstream\_catchment
- merge\_tiles
- read\_geopackage
- reclass\_raster
- report\_no\_data
- set\_no\_data
- snap\_to\_network

# Package ‘hydrographr’

May 25, 2023

**Type** Package

**Title** Scalable Hydrographic Data Processing in R

**Date** 2023

**Version** 1.0.14

**Maintainer** Maria Üblacker <maria.ueblacker@igb-berlin.de> and  
Afroditi Grigoropoulou <afroditi.grigoropoulou@igb-berlin.de>

**Description** Scalable hydrographic geospatial data processing tools using open-source command-line utilities. The package provides functions to download the Hydrography90m data (<https://essd.copernicus.org/articles/14/4525/2022/>), processing, reading and extracting information, as well as assessing network distances and network connectivity. While the functions are, by default, tailored towards the Hydrography90m data, they can also be generalized towards other data and purposes, such as efficient cropping and merging of raster and vector data, point-raster extraction, raster reclassification, and data aggregation. The package depends on the open-source software GDAL/OGR, GRASS-GIS and the AWK programming language in the Linux environment, allowing a seamless language integration. Since the data is processed outside R, hydrographr allows creating scalable geo-processing workflows. Please see the installation guide of the additional software at <https://glowabio.github.io/hydrographr/articles/hydrographr.html>. Windows users need to first activate the Windows Subsystem for Linux (WSL) feature.

**License** GPL-3

**URL** <https://glowabio.github.io/hydrographr/>

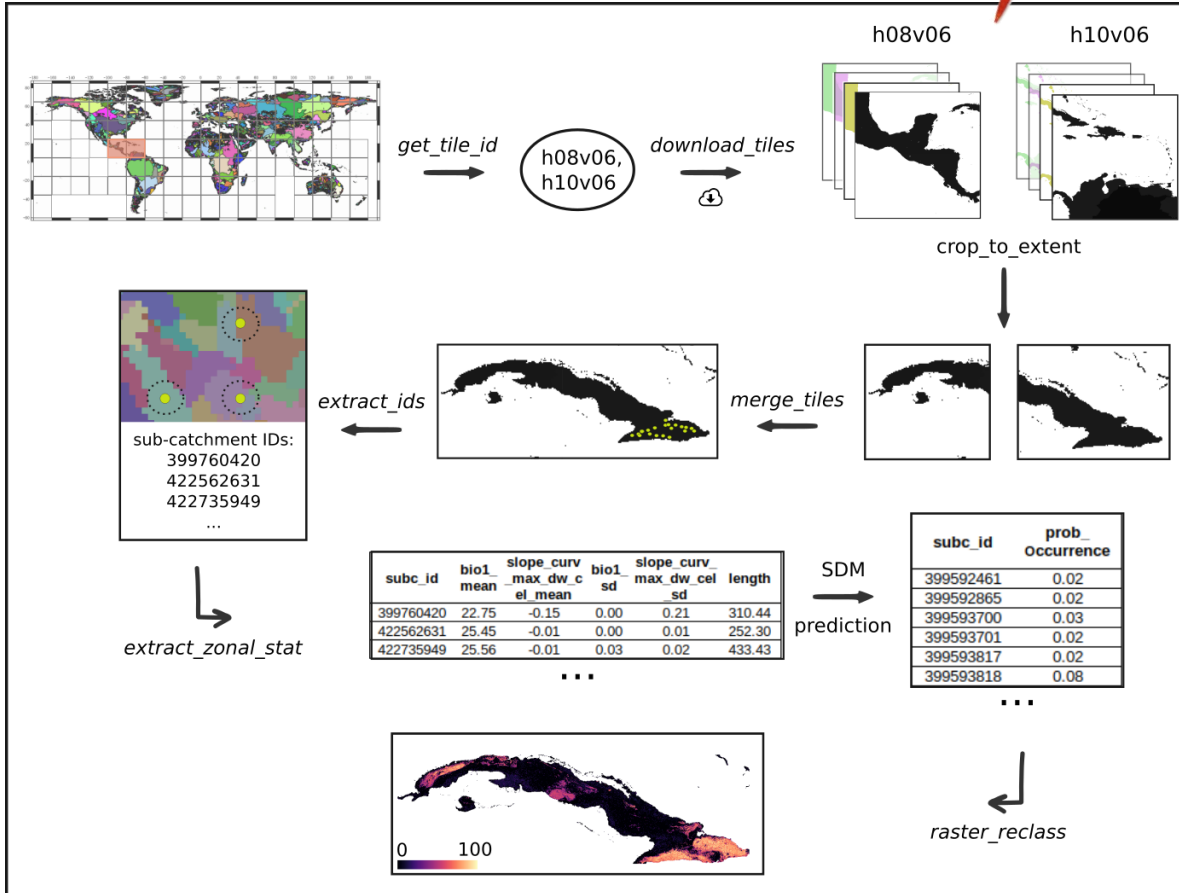
**BugReports** <https://github.com/glowabio/hydrographr/issues>

**Encoding** UTF-8

**LazyData** true

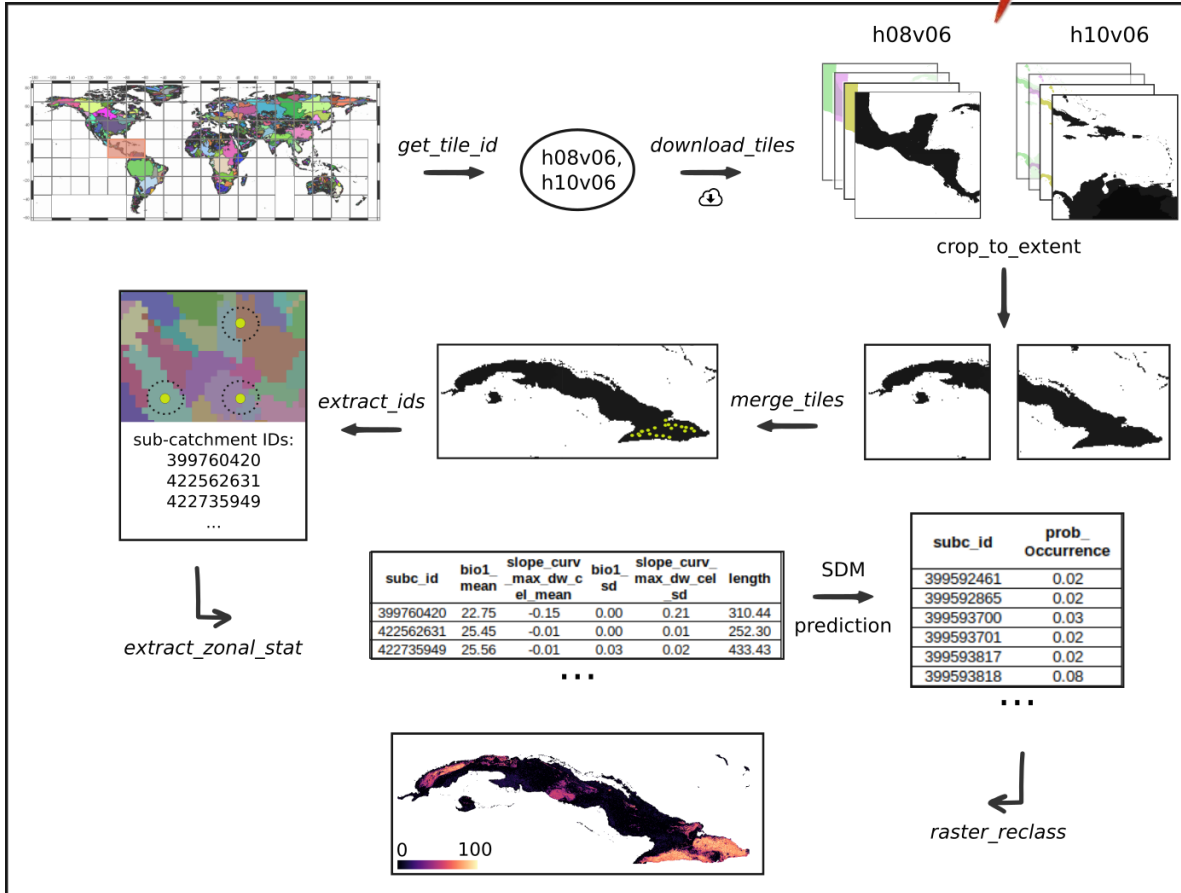
# hydrographr vignettes – example workflows

## Species distribution modelling



# hydrograhr vignettes – example workflows

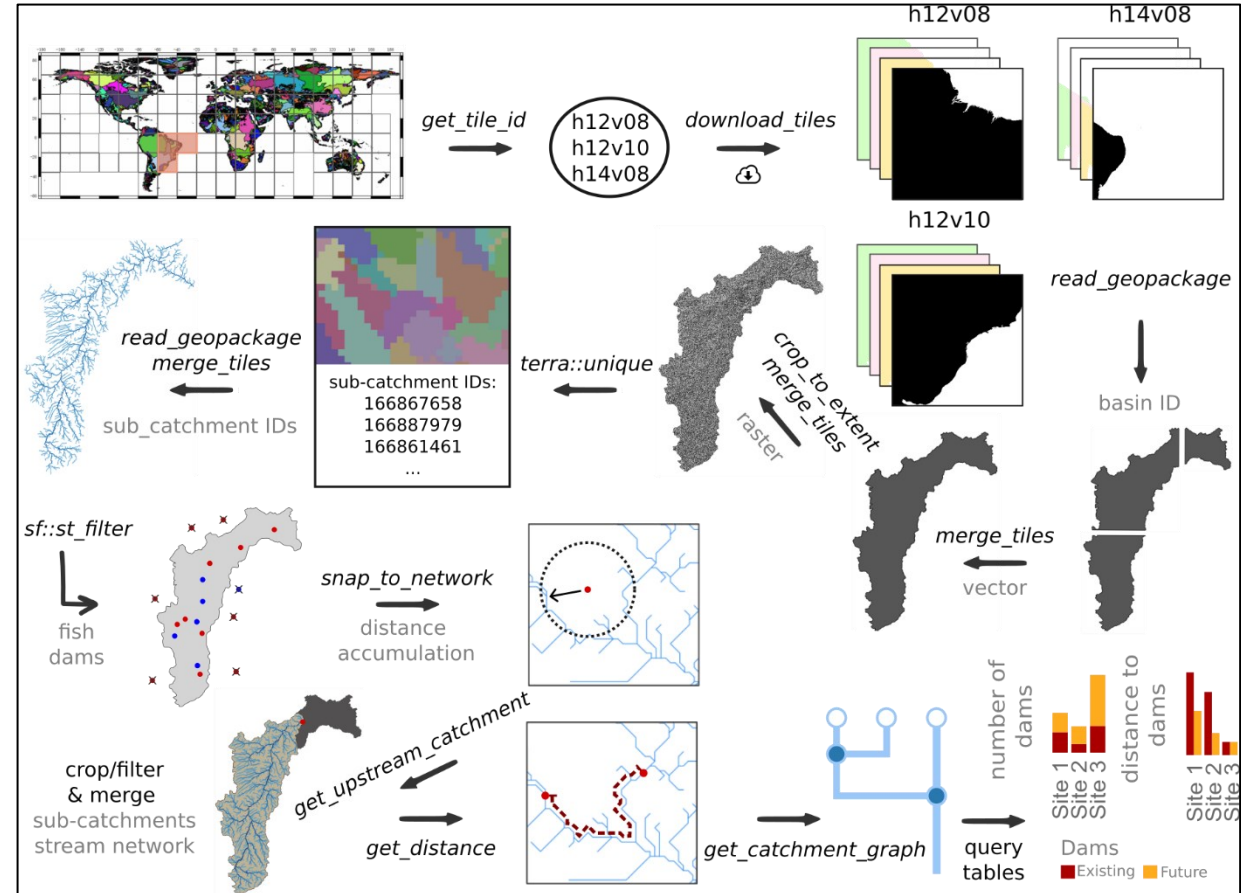
## Species distribution modelling



## Free-flowing rivers

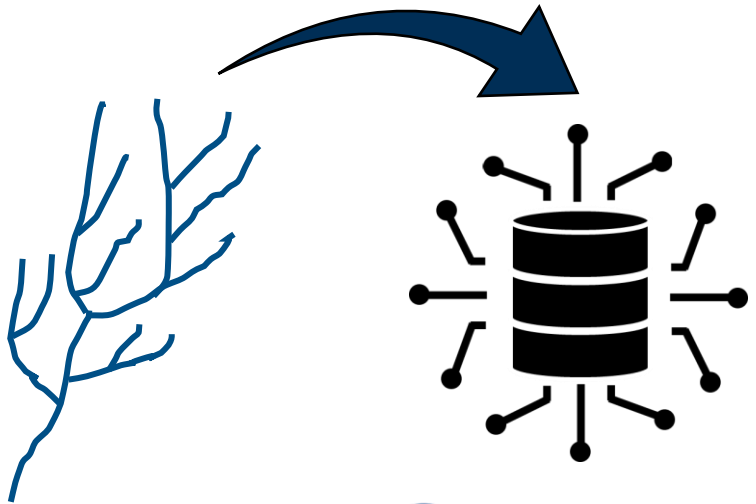


*Conorhynchos conirostris*  
© Paulo Henrique Fiote

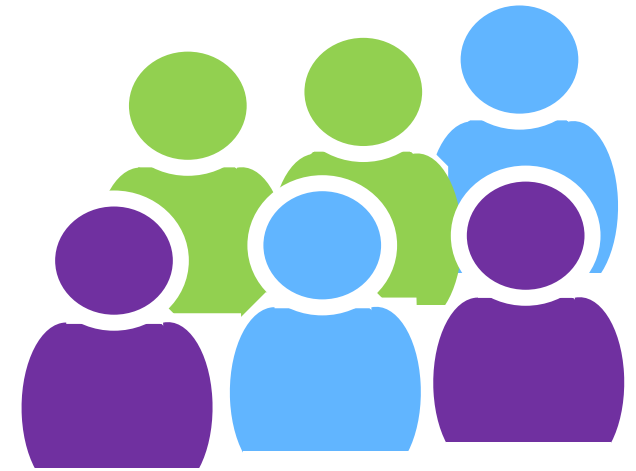


# New online platform – geofresh.org

## Scalable database



## Scientific communities



<https://geofresh.org/>

Allows

- spatial queries
- filtering
- aggregation
- downloading



1. Upload point coordinates



2. Snap points to stream network



### Upload and snap

Please provide your point data as a .csv table with three columns: a unique 'id', 'longitude', 'latitude'.

Coordinates should be provided in the WGS84 coordinate reference system. Column names are flexible. The number of points in your .csv file is currently limited to 1000, and upload file size should not exceed 1MB.

**Point data (.csv format)**

Browse... No file selected

**Load test data**

**Snapping method: sub-catchment (default)**  
Points will be snapped to the nearest location on the river segment of the sub-catchment the point falls in.

**Snap points**

100%

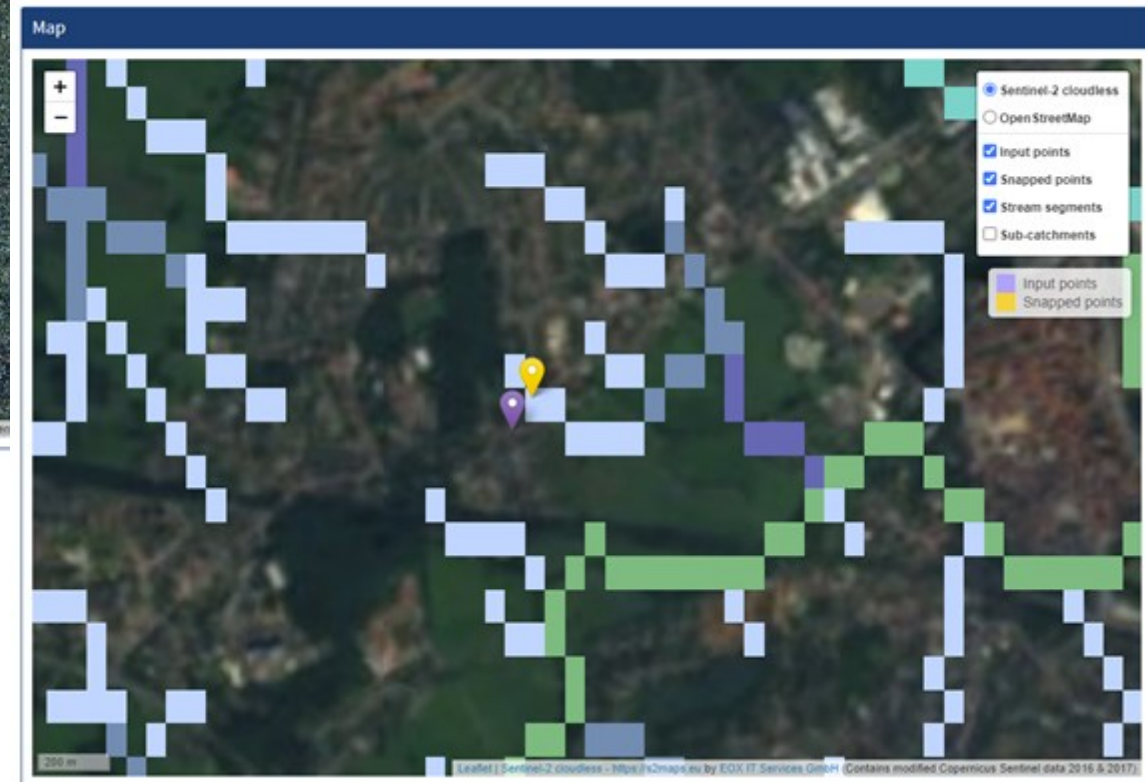
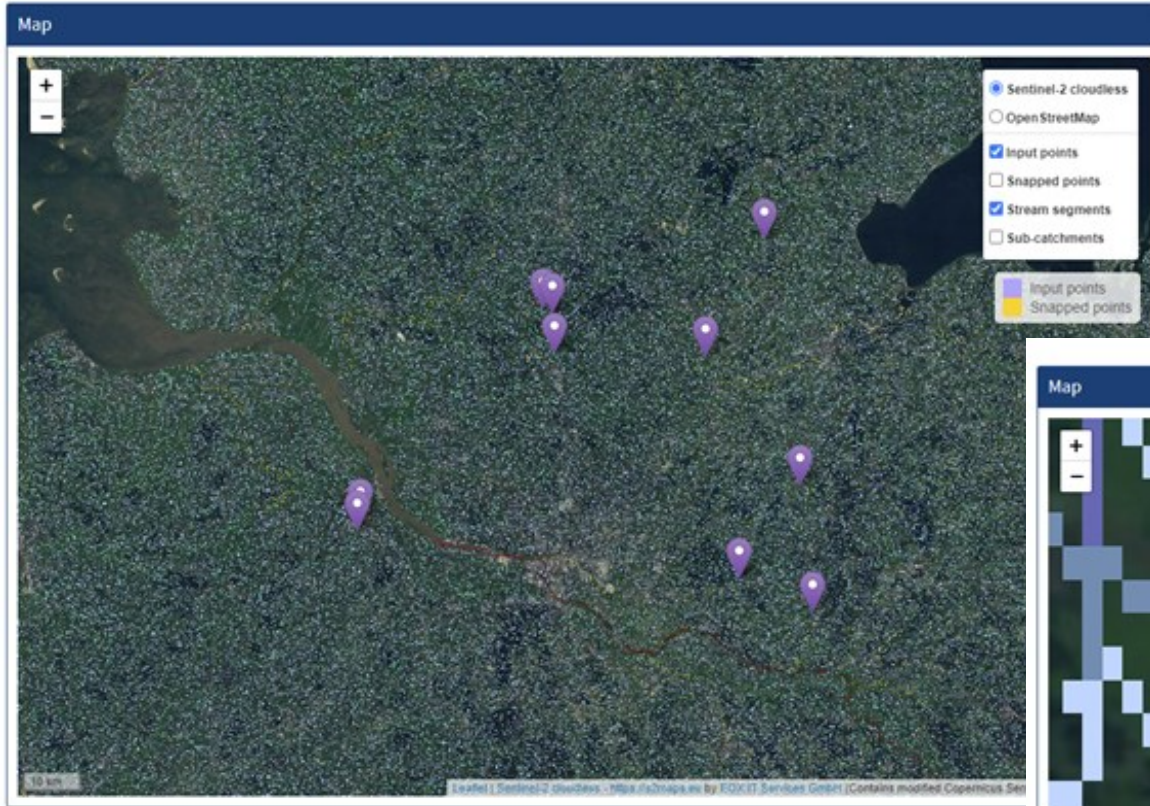
Show **10** entries Search:

ID	longitude	latitude	longitude_snap	latitude_snap	sub-catchment_ID
4058613271	9.9609	53.8553	9.961133	53.855533	506463635
4058539759	10.6366	53.4573	10.63625	53.457917	506627827
4058538795	9.9576	53.9178	9.9576	53.917917	506440890
4058558992	9.4446	53.5823	9.445417	53.582917	506571515

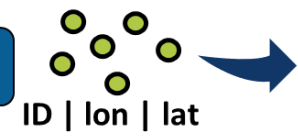
Showing 1 to 10 of 10 entries Previous **1** Next

**Download**

# geofresh.org



1. Upload point coordinates



2. Snap points to stream network



3. Visualize points on the map



1. Upload point coordinates



2. Snap points to stream network



3. Visualize points on the map



4. Select environmental variables

- Topography
- Climate
- Soil
- Land cover

Select environmental variables

Activate the checkboxes to select the required environmental information that should be summarized within the upstream catchment of each point. Please see the source and the citation for each category under the 'Documentation' tab.

Topography	Climate	Soil	Landcover
<input type="checkbox"/> Mean elevation (elev)	<input type="checkbox"/> Annual mean temperature (bio1)	<input type="checkbox"/> Derived saturated water content (awct5)	<input type="checkbox"/> Cropland, rainfed (c10)
<input type="checkbox"/> Flow accumulation (flowpos)	<input type="checkbox"/> Mean diurnal range (bio2)	<input type="checkbox"/> Clay content (clyppt)	<input type="checkbox"/> Cropland, irrigated/post-flooding (c20)
<input type="checkbox"/> Cell maximum curvature	<input type="checkbox"/> Isothermality (bio3)	<input type="checkbox"/> Sand content (sndppt)	<input type="checkbox"/> Cropland/natural vegetation (c30)
<input type="checkbox"/> Cell minimum curvature	<input type="checkbox"/> Temperature seasonality (bio4)	<input type="checkbox"/> Silt content (sltppt)	<input type="checkbox"/> Natural vegetation/cropland (c40)
<input type="checkbox"/> Cell elevation difference	<input type="checkbox"/> Max temperature of warmest month (bio5)	<input type="checkbox"/> Derived available soil water capacity (wwp)	<input type="checkbox"/> Tree cover, broadleaved, evergreen (c50)
<input type="checkbox"/> Cell gradient	<input type="checkbox"/> Min temperature of coldest month (bio6)	<input type="checkbox"/> Soil organic carbon content (orcdrc)	<input type="checkbox"/> Tree cover, broadleaved, deciduous (c60)
<input type="checkbox"/> Shortest distance to drainage divide	<input type="checkbox"/> Temperature annual range (bio7)	<input type="checkbox"/> Soil pH x 10 in H2O (phihox)	<input type="checkbox"/> Tree cover, needleleaved, evergreen (c70)
<input type="checkbox"/> Longest distance to drainage divide	<input type="checkbox"/> Mean temperature of wettest quarter (bio8)	<input type="checkbox"/> Bulk density (bidfie)	<input type="checkbox"/> Tree cover, needleleaved, deciduous (c80)
<input type="checkbox"/> Nearest downstream stream grid cell	<input type="checkbox"/> Mean temperature of driest quarter (bio9)	<input type="checkbox"/> Cation exchange capacity (cecsol)	<input type="checkbox"/> Tree cover, mixed leaf type (c90)
<input type="checkbox"/> Outlet grid cell in the network	<input type="checkbox"/> Mean temperature of warmest quarter (bio10)	<input type="checkbox"/> Coarse fragments volumetric (crfvol)	<input type="checkbox"/> Tree and shrub (c100)
<input type="checkbox"/> Downstream stream node grid cell	<input type="checkbox"/> Mean temperature of coldest quarter (bio11)	<input type="checkbox"/> Grade of a sub-soil being acid (acdwrp)	<input type="checkbox"/> Herbaceous/tree and shrub (c110)
<input type="checkbox"/> Euclidean distance	<input type="checkbox"/> Annual precipitation (bio12)	<input type="checkbox"/> Depth to bedrock (R horizon) (bdrim)	<input type="checkbox"/> Shrubland (c120)
<input type="checkbox"/> Shortest path	<input type="checkbox"/> Precipitation of wettest month (bio13)	<input type="checkbox"/> Probability of occurrence of R horizon (bdrlog)	<input type="checkbox"/> Grassland (c130)
<input type="checkbox"/> Longest path	<input type="checkbox"/> Precipitation of driest month (bio14)	<input type="checkbox"/> Cumulative probability of organic soil (histpr)	<input type="checkbox"/> Lichens, mosses (c140)
<input type="checkbox"/> Nearest downstream stream pixel	<input type="checkbox"/> Precipitation seasonality (bio15)	<input type="checkbox"/> Sodic soil grade (slgwrp)	<input type="checkbox"/> Sparse vegetation (c150)
<input type="checkbox"/> Outlet grid cell in the network	<input type="checkbox"/> Precipitation of wettest quarter (bio16)		<input type="checkbox"/> Tree cover, flooded, fresh/brackish water (c160)
<input type="checkbox"/> Downstream stream node grid cell	<input type="checkbox"/> Precipitation of driest quarter (bio17)		<input type="checkbox"/> Tree cover, flooded, saline water (c170)
<input type="checkbox"/> Segment downstream mean gradient	<input type="checkbox"/> Precipitation of warmest quarter (bio18)		<input type="checkbox"/> Shrub or herbaceous (c180)
<input type="checkbox"/> Segment upstream mean gradient	<input type="checkbox"/> Precipitation of coldest quarter (bio19)		<input type="checkbox"/> Urban areas (c190)
<input type="checkbox"/> Cell upstream gradient			<input type="checkbox"/> Bare areas (c200)
<input type="checkbox"/> Cell stream course curvature			<input type="checkbox"/> Water bodies (c210)
<input type="checkbox"/> Segment downstream elevation difference			<input type="checkbox"/> Snow and ice (c220)
<input type="checkbox"/> Segment upstream elevation difference			
<input type="checkbox"/> Cell upstream elevation difference			
<input type="checkbox"/> Cell downstream elevation difference			
<input type="checkbox"/> Segment downstream distance			
<input type="checkbox"/> Segment upstream distance			
<input type="checkbox"/> Cell upstream distance			
<input type="checkbox"/> Strahler's stream order (stream_strahler)			
<input type="checkbox"/> Shreve's stream magnitude (stream_shreve)			
<input type="checkbox"/> Horton's stream order (stream_horton)			
<input type="checkbox"/> Hack's stream order (stream_hack)			
<input type="checkbox"/> Topological dimension of streams			
<input type="checkbox"/> Length of the stream reach (length)			
<input type="checkbox"/> Straight length (stright)			
<input type="checkbox"/> Sinusoid of the stream reach (sinusoid)			

# geofresh.org

Start query Query selected environmental variables for the local sub-catchment (min, max, mean, sd)

Selected variables:  
 Topography: elev  
 Climate: bio1  
 Soil: awcts  
 Land cover: c10

Local sub-catchment  
 Topography Climate Soil Land cover  
 Show 10 entries Search:

id	subc_id	elev_min	elev_max	elev_mean	elev_sd
4058539759	506627827	9.4	13.9	11.2938	1.0192
4058560090	506465118	30.8	40.4	35.7722	2.6857
4058623136	506561224	1.9	33	13.5313	9.5675
4058613271	506463635	17.5	19.6	18.3308	0.6591

Showing 1 to 10 of 10 entries Previous 1 Next

Download topography data for local catchment

Plot results

Plot results of environmental variables queries as histograms for topography, climate and soil or boxplots for land cover.

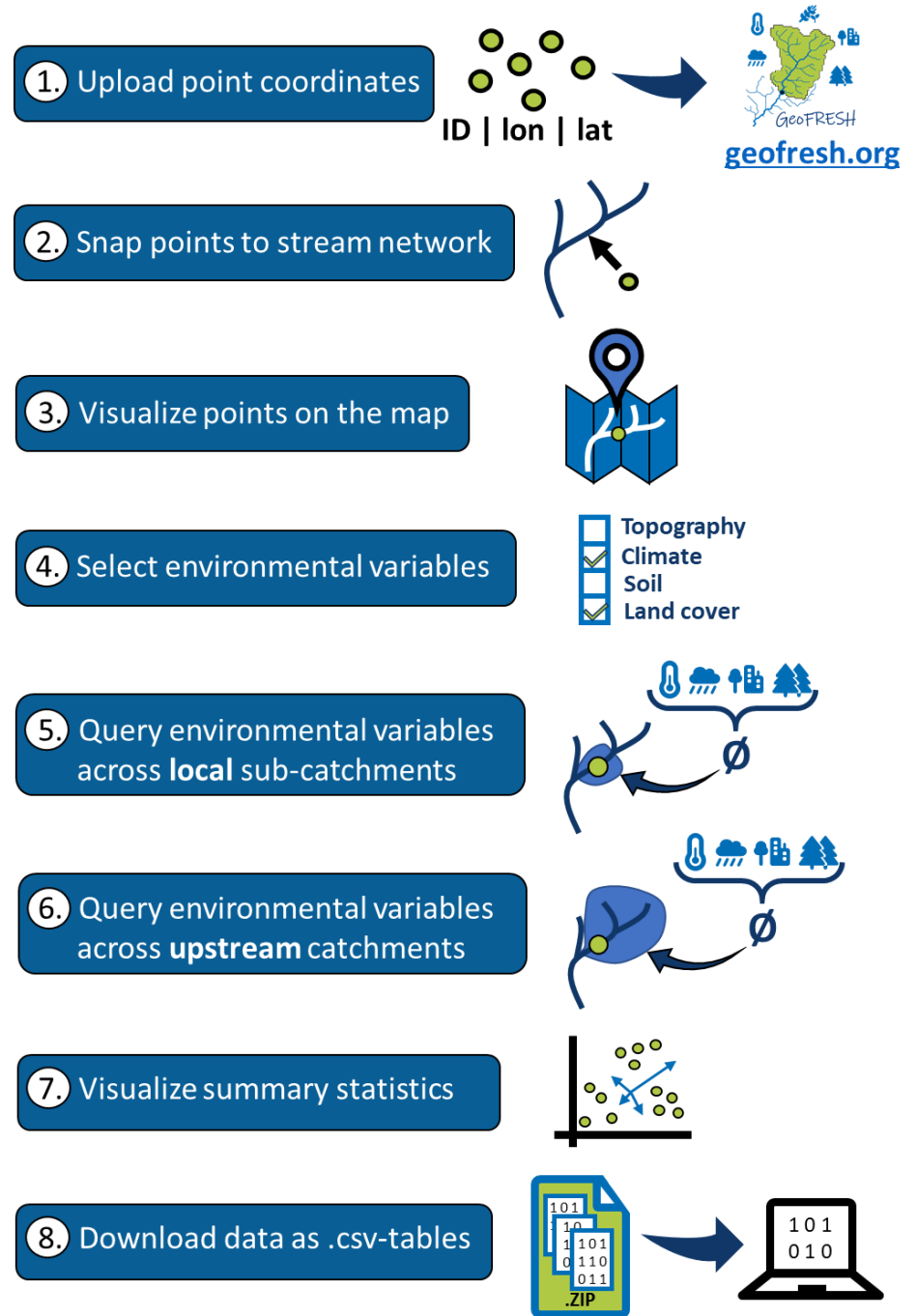
Topography Climate Soil Land cover

Select variable to plot:  
 Annual mean temperature (bio1)

Number of bins:  
 1 10 20

Histogram of "Annual mean temperature (bio1)" for local sub-catchment

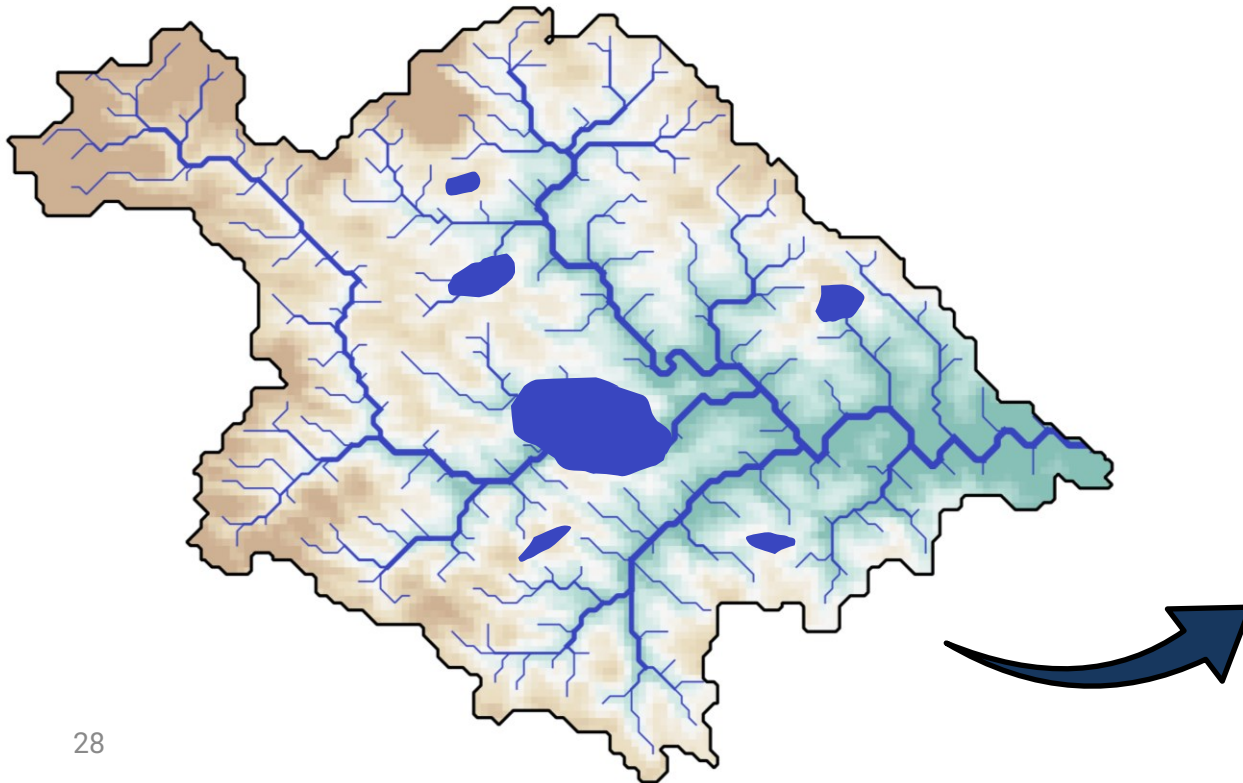
Histogram of "Annual mean temperature (bio1)" for upstream catchment





# New online platform – geofresh.org

- Connectivity as a default instead of a feature
- Address standing water bodies (NASA SWOT data)
- Research reproducibility (Virtual Research Environments)



GeoFRESH

Home Analysis Tutorial Documentation

Welcome to the GeoFRESH platform

GeoFRESH is a platform that helps freshwater researchers to process point data across the global river network by providing a set of spatial tools.

Follow the demo or upload your csv-table that contains geographic coordinates. GeoFRESH allows you to

- map your points,
- move points to the nearest stream network segment,
- delineate upstream catchments of each point,
- extract a suite of environmental attributes across the catchment,
- and download the data for further analyses.

GeoFRESH is based on the Hydrography90m stream network. For more information, please see the publication and hydrography.org regarding the single data layers.

Analysis  
Check out the different analysis steps.

Tutorial  
Check out the full functionality of the platform.

Documentation  
Learn about the project background.

NFDI4Earth

GeoFRESH was funded by NFDI4Earth and the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB).

IGB  
Leibniz Institute of Freshwater Ecology and Inland Fisheries

Privacy Policy



ALL RESOURCES ▾ AquaINFRA Search 🔍

### 1.654 SEARCH RESULTS

- Articles & Reports** | Mediterranean region case study

[Phenological sensitivity and seasonal variability explain climate-driven trends in Mediterranean butterflies](#)

Colom, Pau et al.

Although climate-driven phenological shifts have been...

Inspect

Added
- Services** | Baltic Sea region case study

[Indicators and assessments WMS](#)

HELCOM Secretariat

This web mapping service contains HELCOM thematic...

Inspect

Add to VRE
- Research data** | North Sea region case study

[High-resolution bathymetry model for the Lena Delta region](#)

Matthias Fuchs et al.

This data set includes spatial high-resolution bathymetry...

Inspect

Added
- Open Educational Resources** | Pan-European Biodiversity case study

[River Culture – Life as a dance to the rhythm of the waters](#)

UNESCO

The global crises of biodiversity and cultural diversity are...

Inspect

Add to VRE
- Software & Source code** | North Sea region case study

[CRITTERBASE, a science-driven data warehouse for marine biota](#)

Katharina Teschke et al.

Data on marine biota exist in many formats and sources...

Inspect

Add to VRE

Spatial coverage ▾



Apply spatial filter

Temporal coverage ▾



Apply temporal filter

Publication date ▾



# Thank you!



NFDI 4  
BIODIVERSITY



Jaime  
Garcia Marquez



Marlene  
Schürz



Afroditi  
Grigoropoulou



Yusdiel  
Torres Cambas



Vanessa  
Bremerich



Christoph  
Schürz



Merret  
Buurman



Thomas  
Tomiczek



Giuseppe  
Amatulli



Longzhu  
Shen



Jens  
Kiesel



Mathieu  
Floury

eOSC | AqualNFRA



Funded by  
the European Union

Yale

SPATIAL-ECOLOGY

# Comparison with the terra package

- merging
- cropping
- re-classification
- zonal statistics
- hydrographr shows high performance
- RAM-efficient processing

