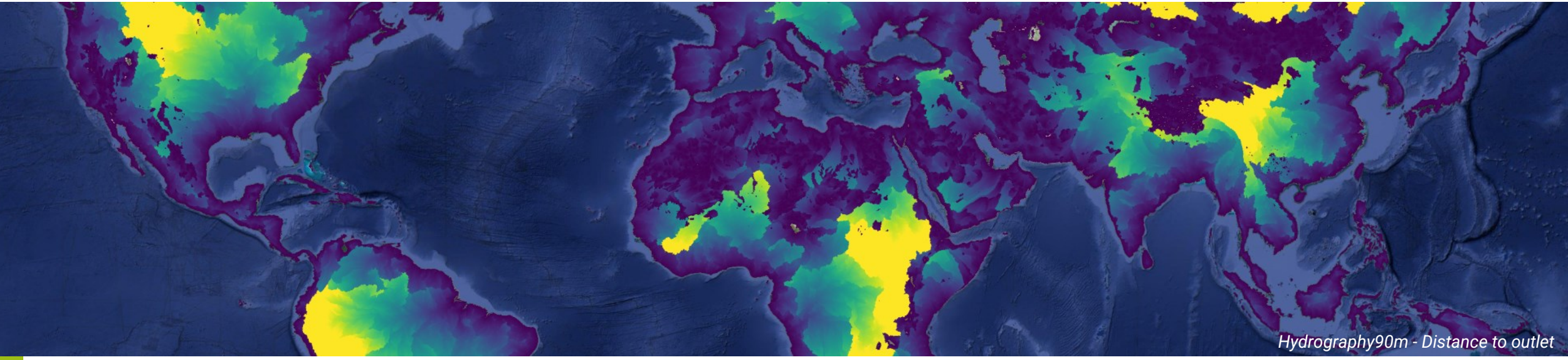


Tools to address the longitudinal connectivity in spatial freshwater biodiversity science

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

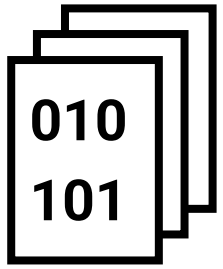
- Complex geo-spatial data processing
- Integration of connectivity and routing
- Large amount of data
- Scalability across large spatio-temporal scales

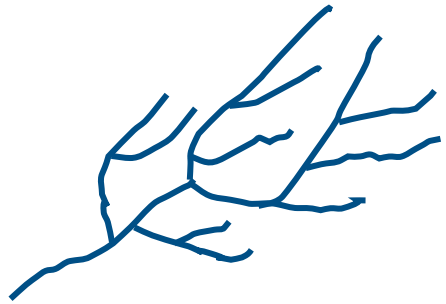
IGB

How to address these limitations?

Support scientific communities with

- **scalable processing workflows**

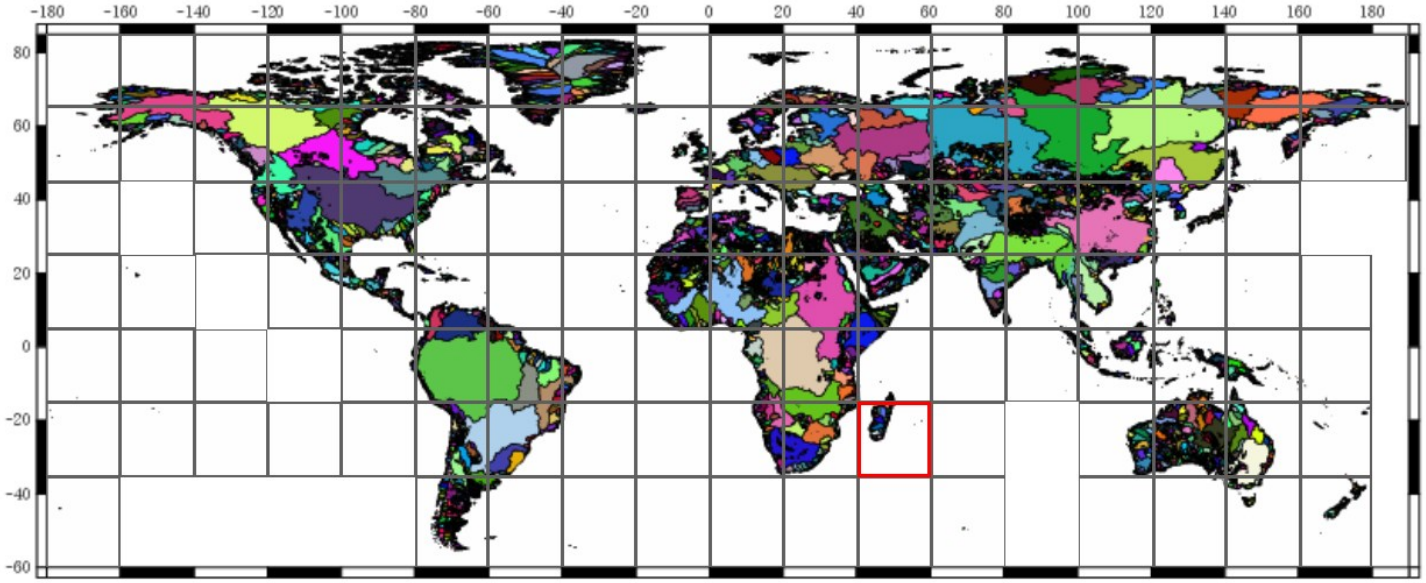




- **easy-to-use-tools for geospatial data processing**

The Hydrography90m dataset

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



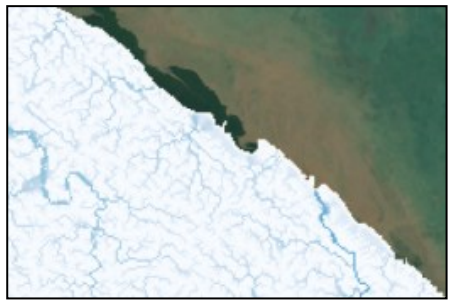
https://hydrography.org/

Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M., Shen, L. & Domisch, S. (2022) Hydrography90m: A new high-resolution global hydrographic dataset. *Earth System Science Data 14(10), 4525-4550*

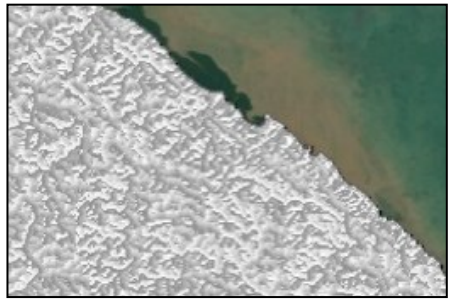
IGB

The Hydrography90m dataset

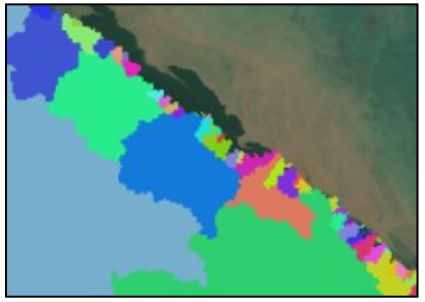
- 1.6 million drainage basins
- 726 million unique stream segments and sub-catchments
- 42 stream topographic variables
- Only open-source tools!



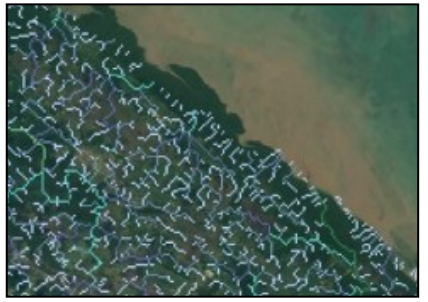
Accumulation



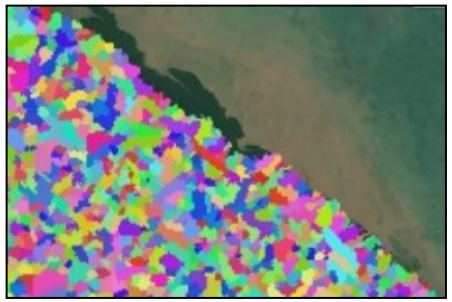
Direction



Basins



Network



Sub-catchments

https://hydrography.org/

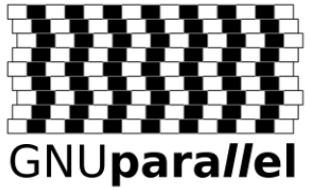
Yale Center for Research Computing

Amatulli, G., Garcia Marquez, J., Sethi, T., Kiesel, J., Grigoropoulou, A., Üblacker, M., Shen, L. & Domisch, S. (2022) Hydrography90m: A new high-resolution global hydrographic dataset. *Earth System Science Data 14(10), 4525-4550*

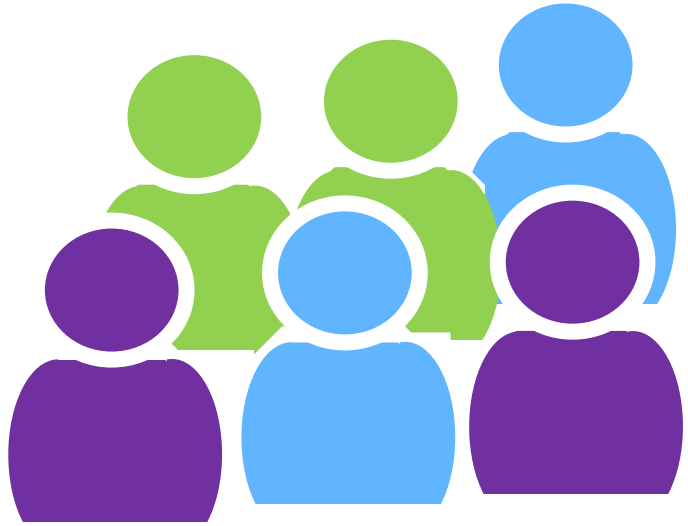
IGB

Open-source toolbox

Scalable, command-line tools

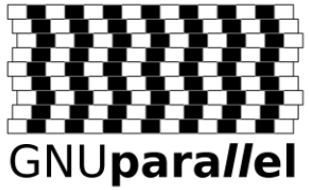


Scientific communities

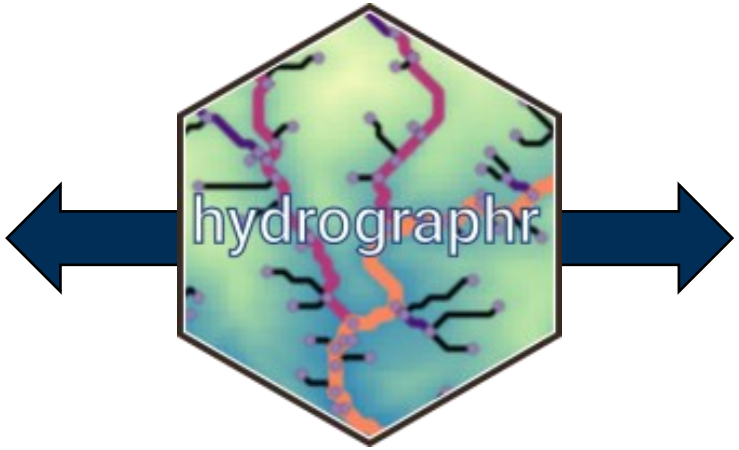


Open-source toolbox

Scalable, command-line tools

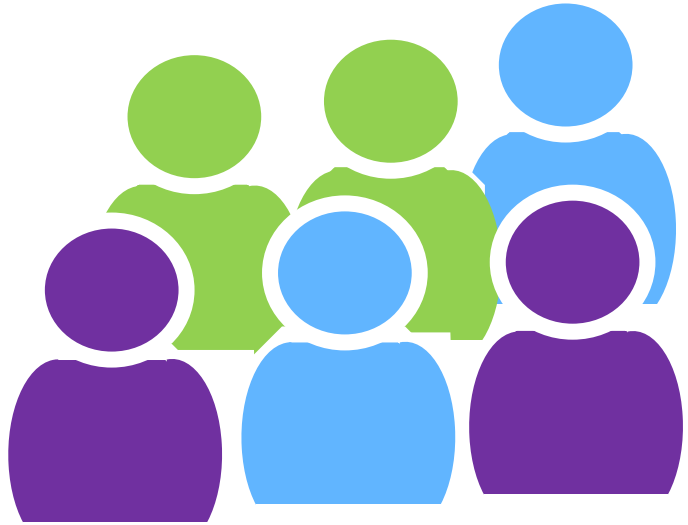


hydrographr



- Wrapper functions
- Scalability
- R language

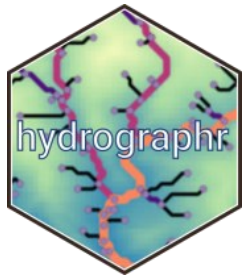
Scientific communities



Üblacker, M., Grigoropoulou,. A., Garcia Marquez, J., Torres Cambas, Y., Tomiczek, T., Floury, M., Bremerich, V., Schürz, C. Amatulli, G., Grossart, H-P., Domisch, S. hydrographr: an R package for scalable hydrographic data processing. Methods in Ecology and Evolution (2023)

IGB

Features of the hydrographr package

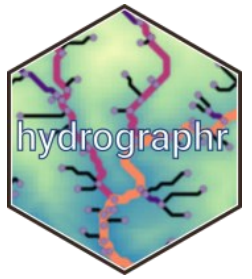


- Supports freshwater geospatial data processing
- Raster, vector, graph, data.table
- Hydrography90m dataset (hydrography.org)
- Data processing outside R --> allows scalable analyses, handles massive data
- Currently 24 functions for
    - downloading
    - data processing
    - reading & extracting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

Üblacker, M., Grigoropoulou,. A., Garcia Marquez, J., Torres Cambas, Y., Tomiczek, T., Floury, M., Bremerich, V., Schürz, C. Amatulli, G., Grossart, H-P., Domisch, S. hydrographr: an R package for scalable hydrographic data processing. Methods in Ecology and Evolution (2023)

IGB

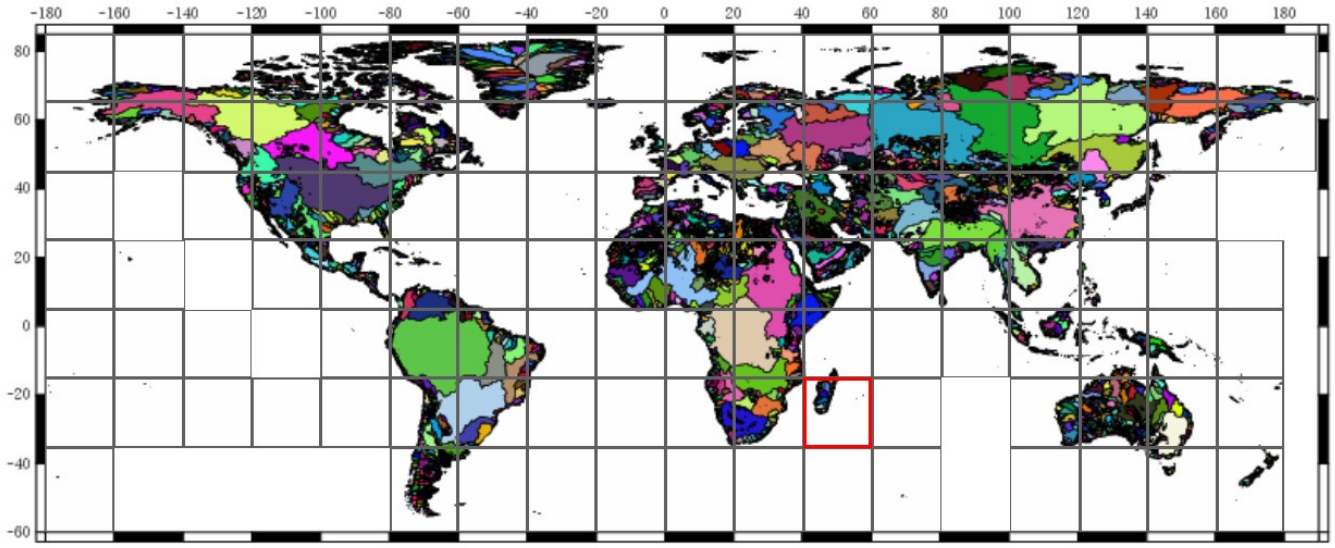
hydrographr functions – Downloading data



- 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()



https://github.com/glowabio/hydrographr

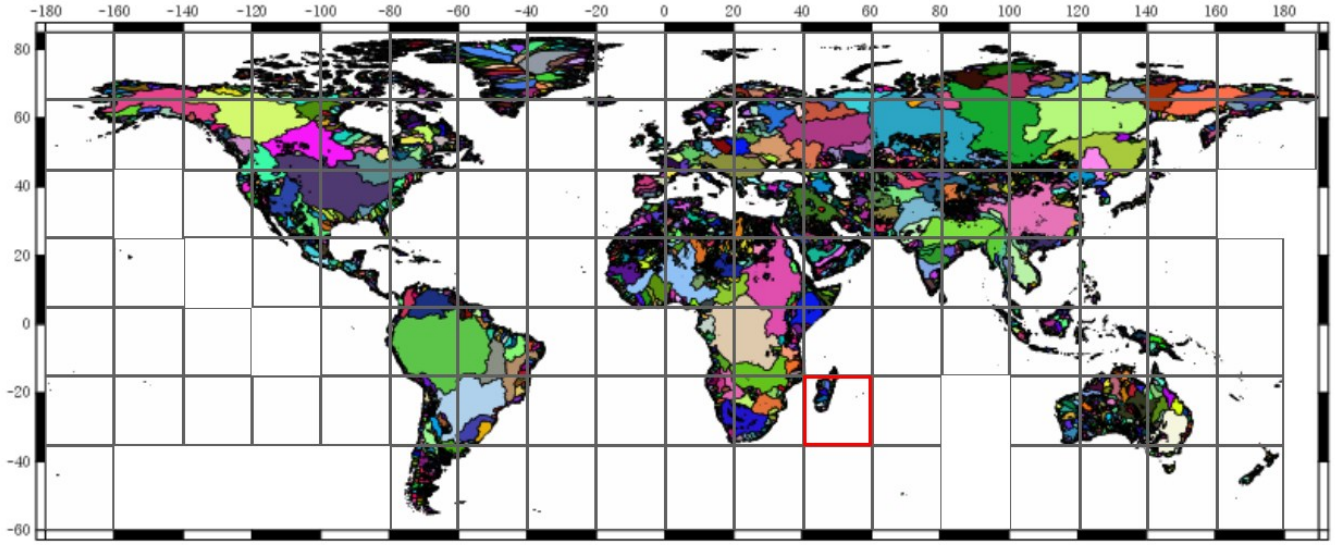
IGB

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()



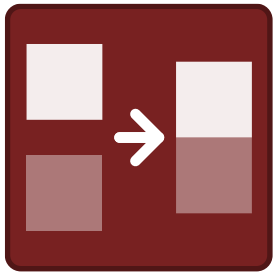
https://github.com/glowabio/hydrographr

IGB

hydrographr functions – Data processing

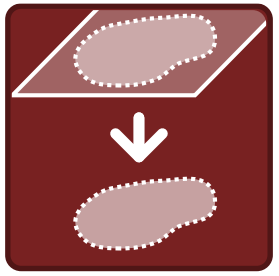
Heavy data processing is done outside R, allowing to address large study areas

merge_tiles()

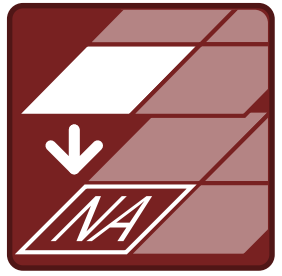


snap_to_subc_segment()

crop_to_extent()

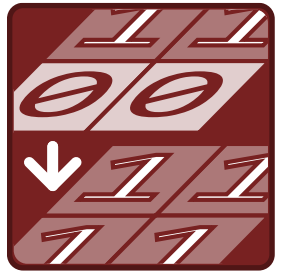


set_no_data()



snap_to_network()

reclass_raster()



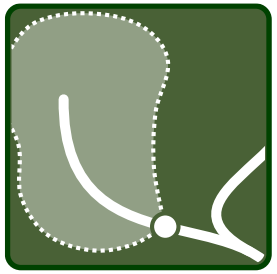
https://github.com/glowabio/hydrographr

IGB

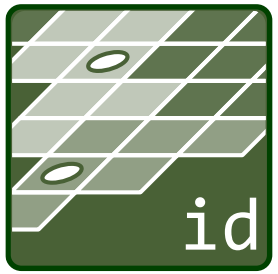
hydrographr functions – Reading & extracting data

Reading & extracting large amounts of data

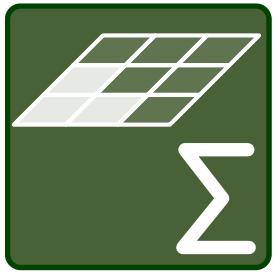
get_upstream_catchment()



extract_ids()



extract_zonal_stat()

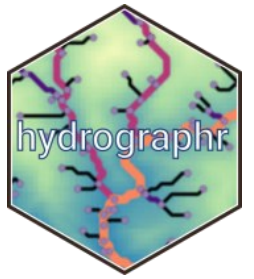


report_no_data()

read_geopackage()

https://github.com/glowabio/hydrographr

hydrographr functions – Distance & connectivity

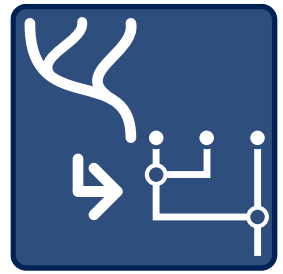


Assess point distances along the network & use graphs to analyse the network

get_distance()

get_catchment_graph()

R igraph



get_segment_neighbours()

R igraph

get_pfafstetter_basins()

R igraph

get_centrality()

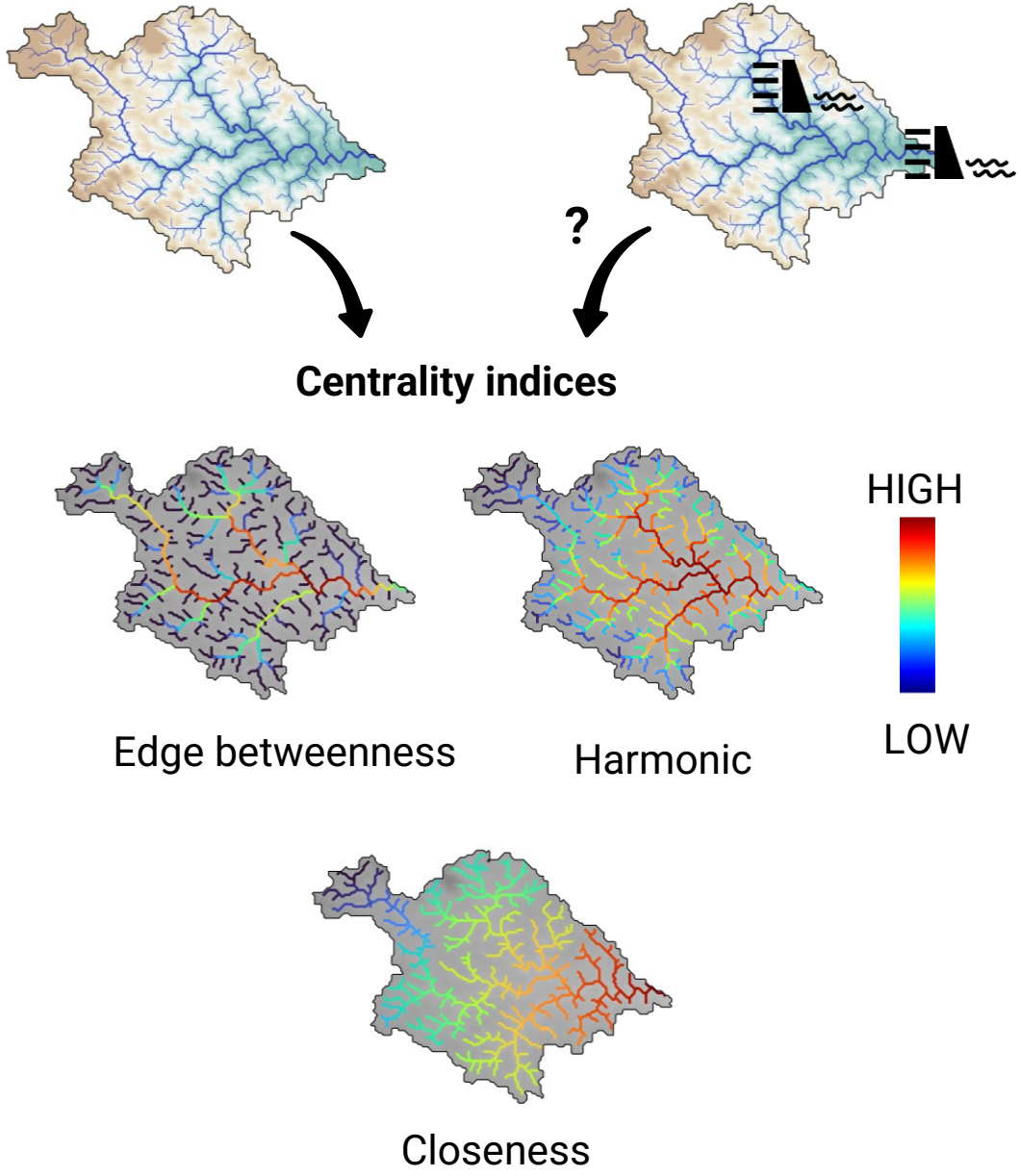
https://github.com/glowabio/hydrographr

IGB

Connectivity & network structure

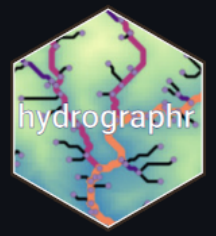
Across large spatial scales, assess the

- 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
- …



README.md

hydrographr



DOI 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!

NFDI 4 BIODIVERSITY

NFDI4Earth

README.md

hydrographr

DOI 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)
```

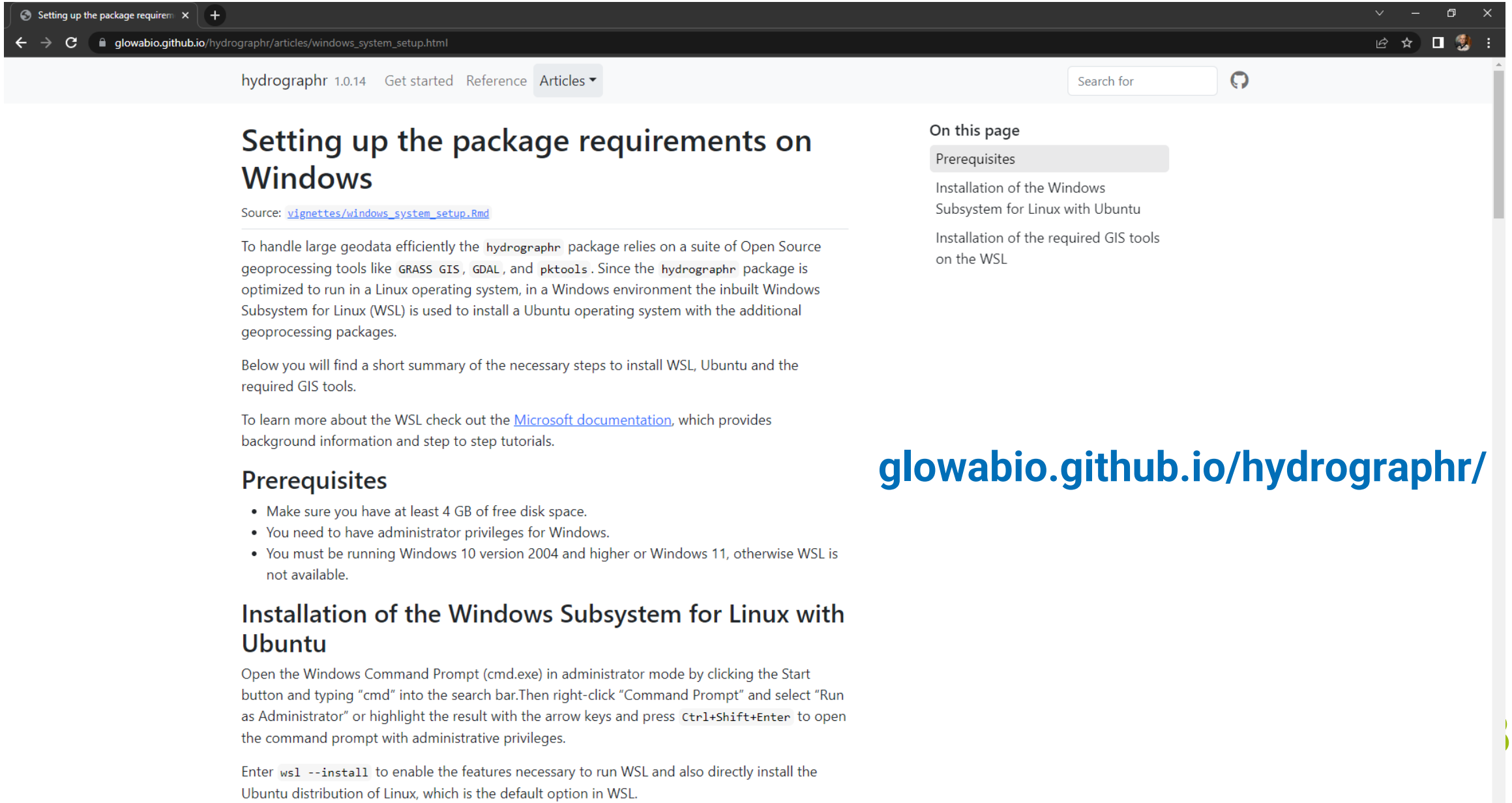
The pdf manual of the `hydrographr` package can be downloaded here.

We thank NFDI4Biodiversity and NFDI4Earth for providing the funding that helped us getting the hydrographr package together!

NFDI 4 BIODIVERSITY

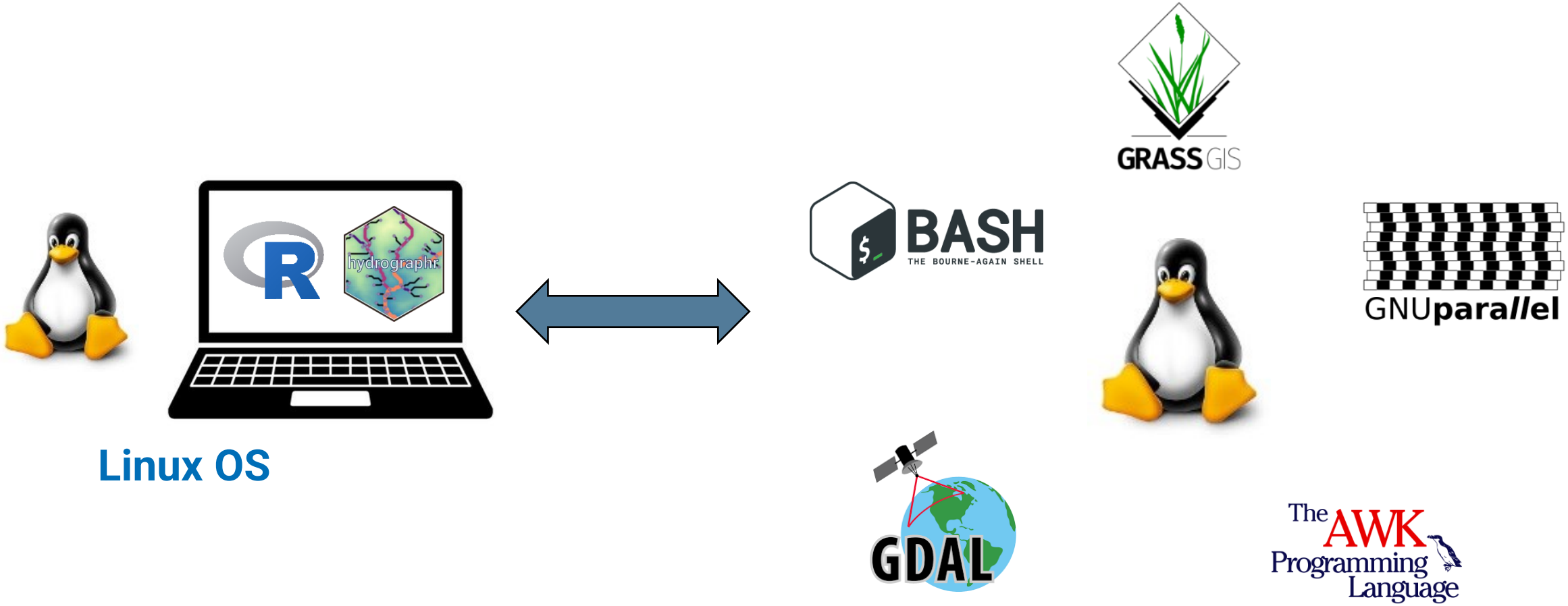
NFDI4Earth

One-time installation of the software



glowabio.github.io/hydrographr/

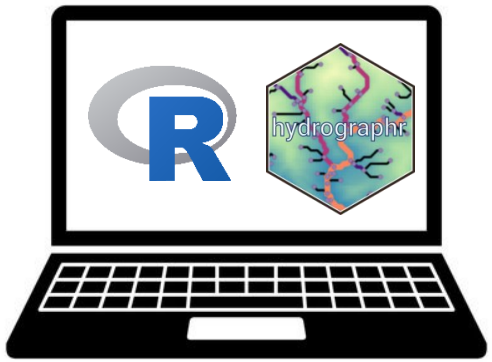
Data processing with open-source tools

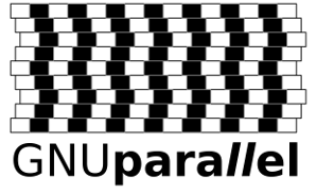


Icon sources: https://linuxiac.com/windows-subsystem-for-linux-explained-wsl-wsl2/; https://de.m.wikipedia.org/wiki/Datei:Windows_logo_-_2012.svg; https://stock.adobe.com/de/search?k=laptop+icon; https://www.creativefabrica.com/de/product/computer-cpu-icon-vector/

IGB

Data processing with open-source tools





Windows OS

Windows Subsystem for Linux (WSL)

Icon sources: https://linuxiac.com/windows-subsystem-for-linux-explained-wsl-wsl2/; https://de.m.wikipedia.org/wiki/Datei:Windows_logo_-_2012.svg; https://stock.adobe.com/de/search?k=laptop+icon; https://www.creativefabrica.com/de/product/computer-cpu-icon-vector/

IGB

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 to first activate the Windows Subsystem for Linux (WSL) feature.

License GPL-3

URL https://glowabio.github.io/hydrographr/

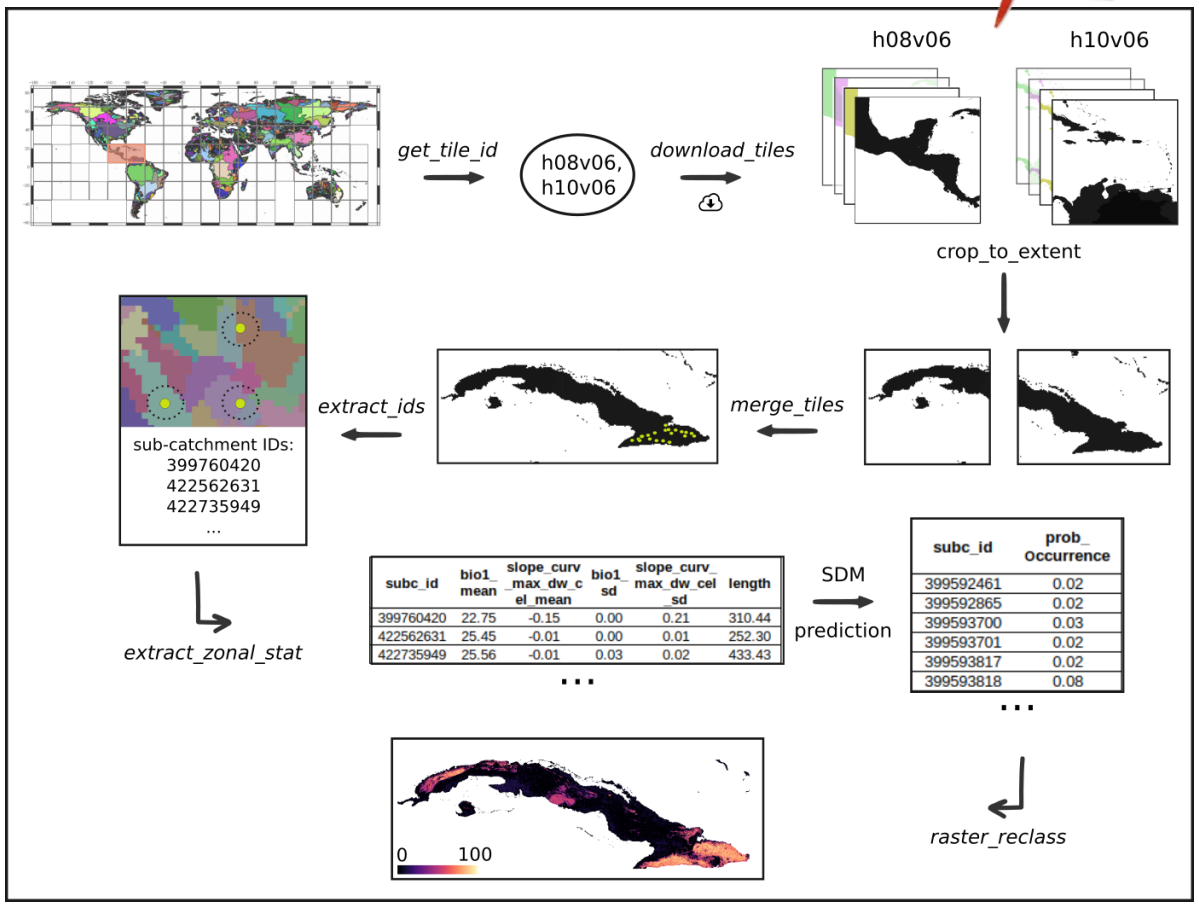
BugReports https://github.com/glowabio/hydrographr/issues

Encoding UTF-8

LazyData true

hydrographr vignettes – example workflows

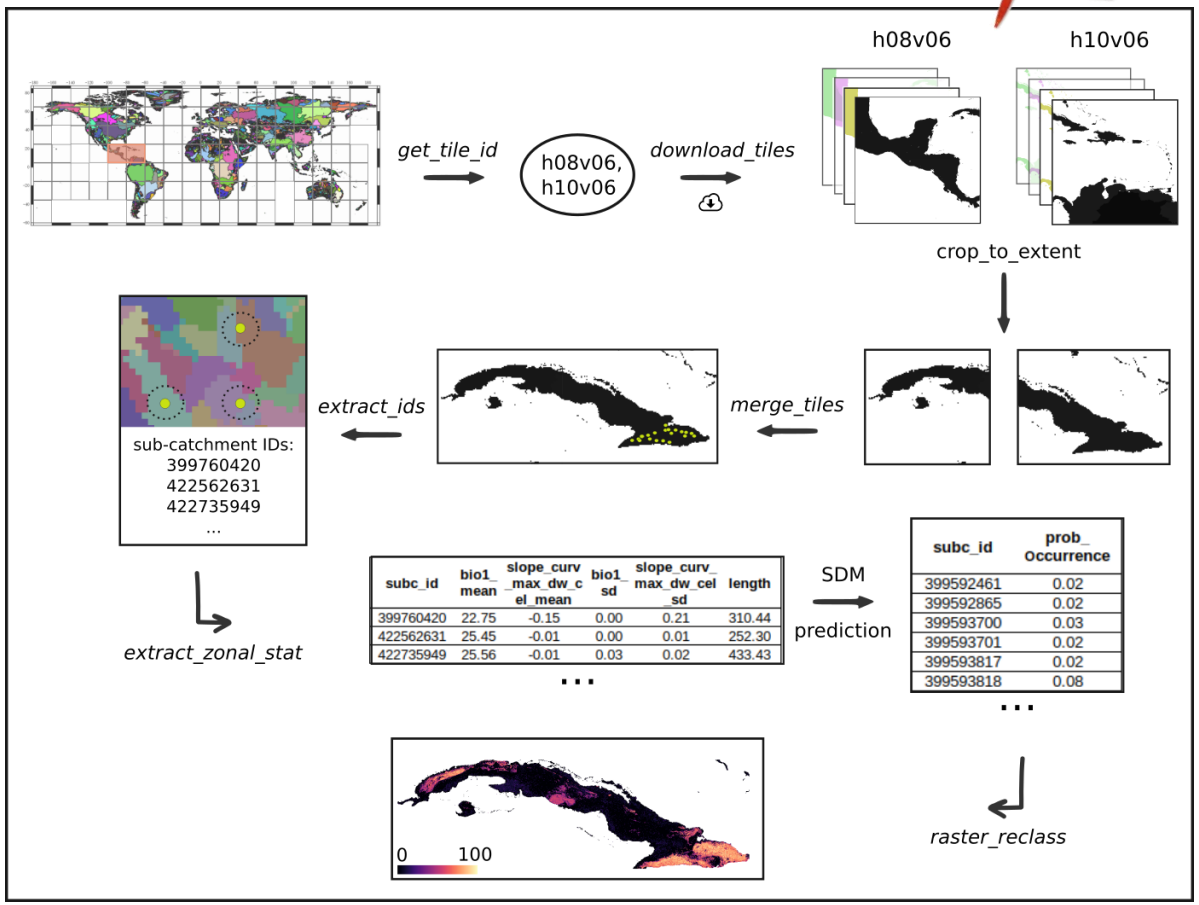
Species distribution modelling



https://glowabio.github.io/hydrographr/

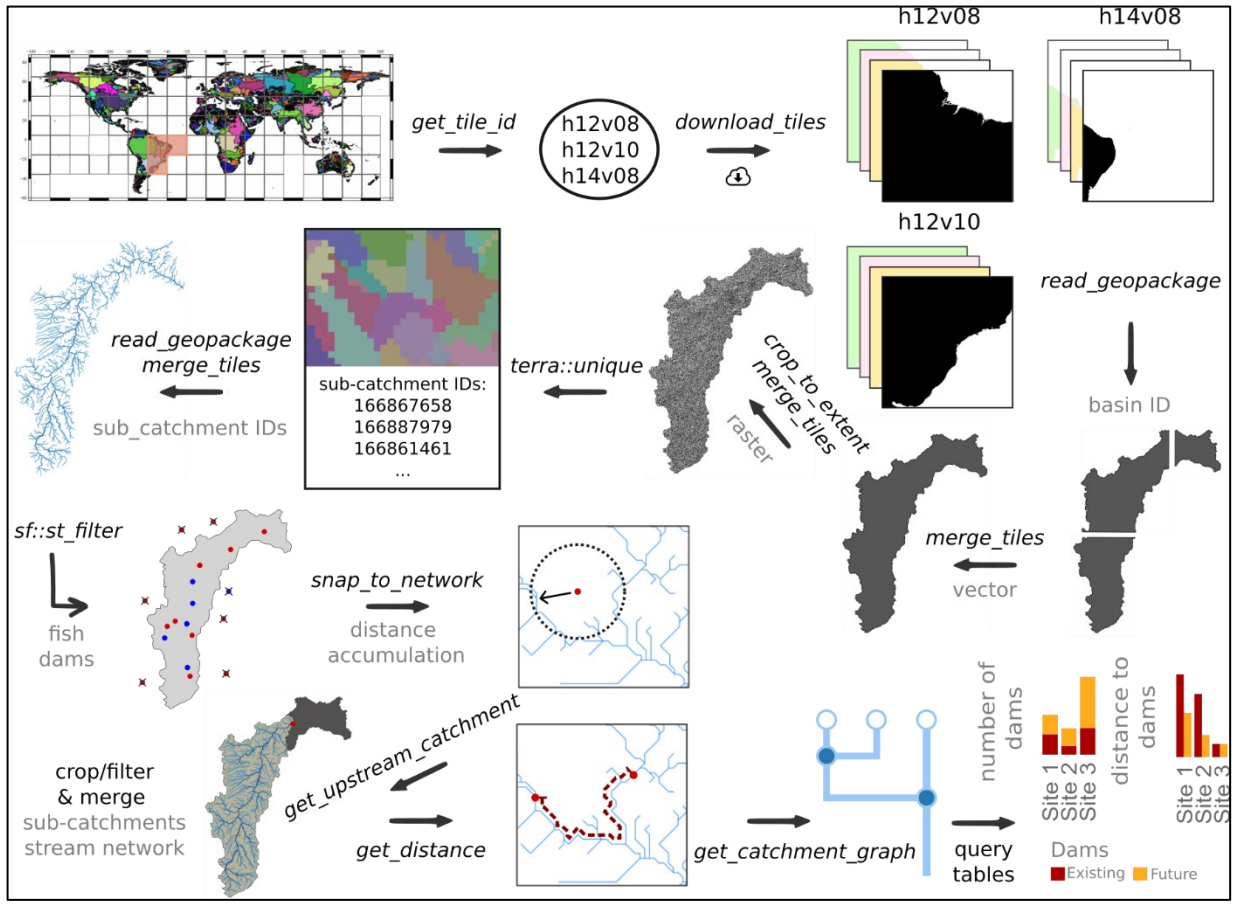
hydrographr vignettes – example workflows

Species distribution modelling



Free-flowing rivers

Conorhynchos conirostris
© Paulo Henrique Fiote

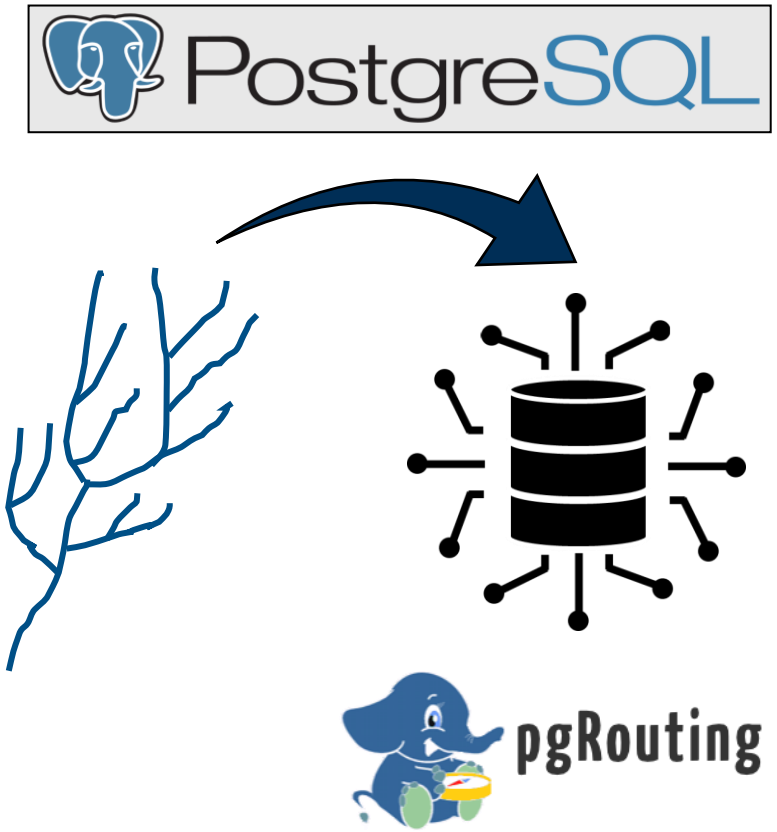


https://glowabio.github.io/hydrographr/

IGB

New online platform – geofresh.org

Scalable database

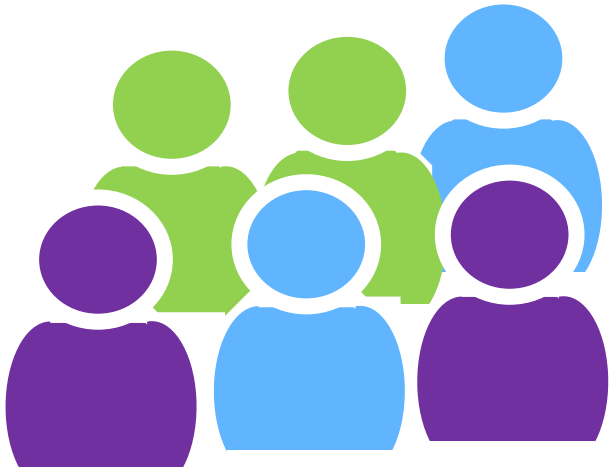


https://geofresh.org/

Allows

- spatial queries
- filtering
- aggregation
- downloading

Scientific communities



IGB

geofresh.org

1. Upload point coordinates

ID | lon | lat

2. Snap points to stream network

geofresh.org

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

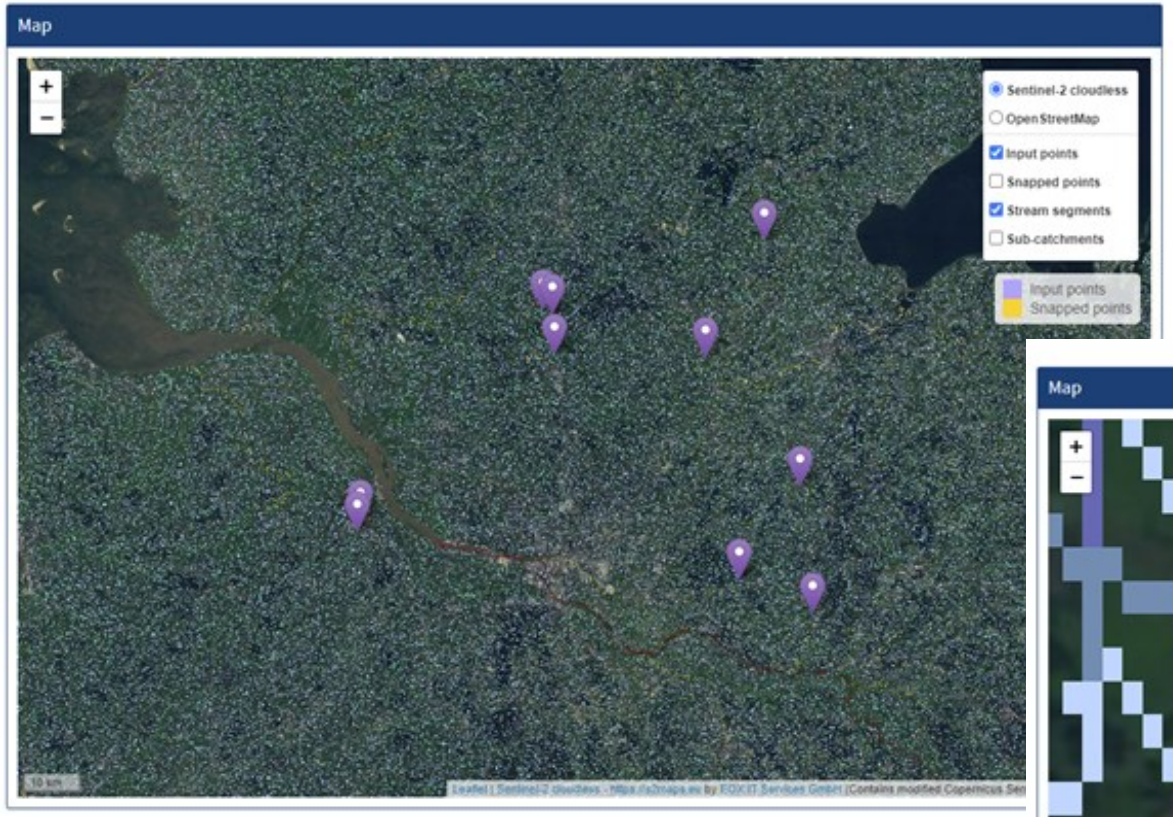
ID | lon | lat

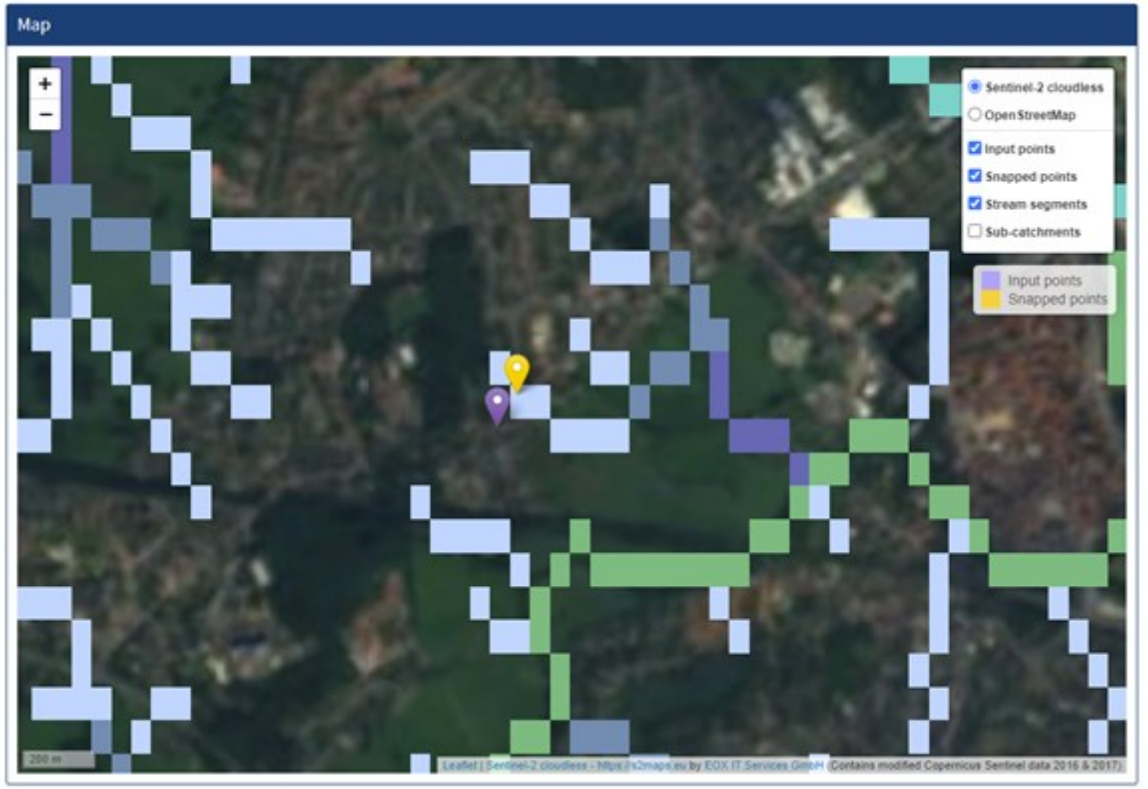
GeoFRESH

geofresh.org

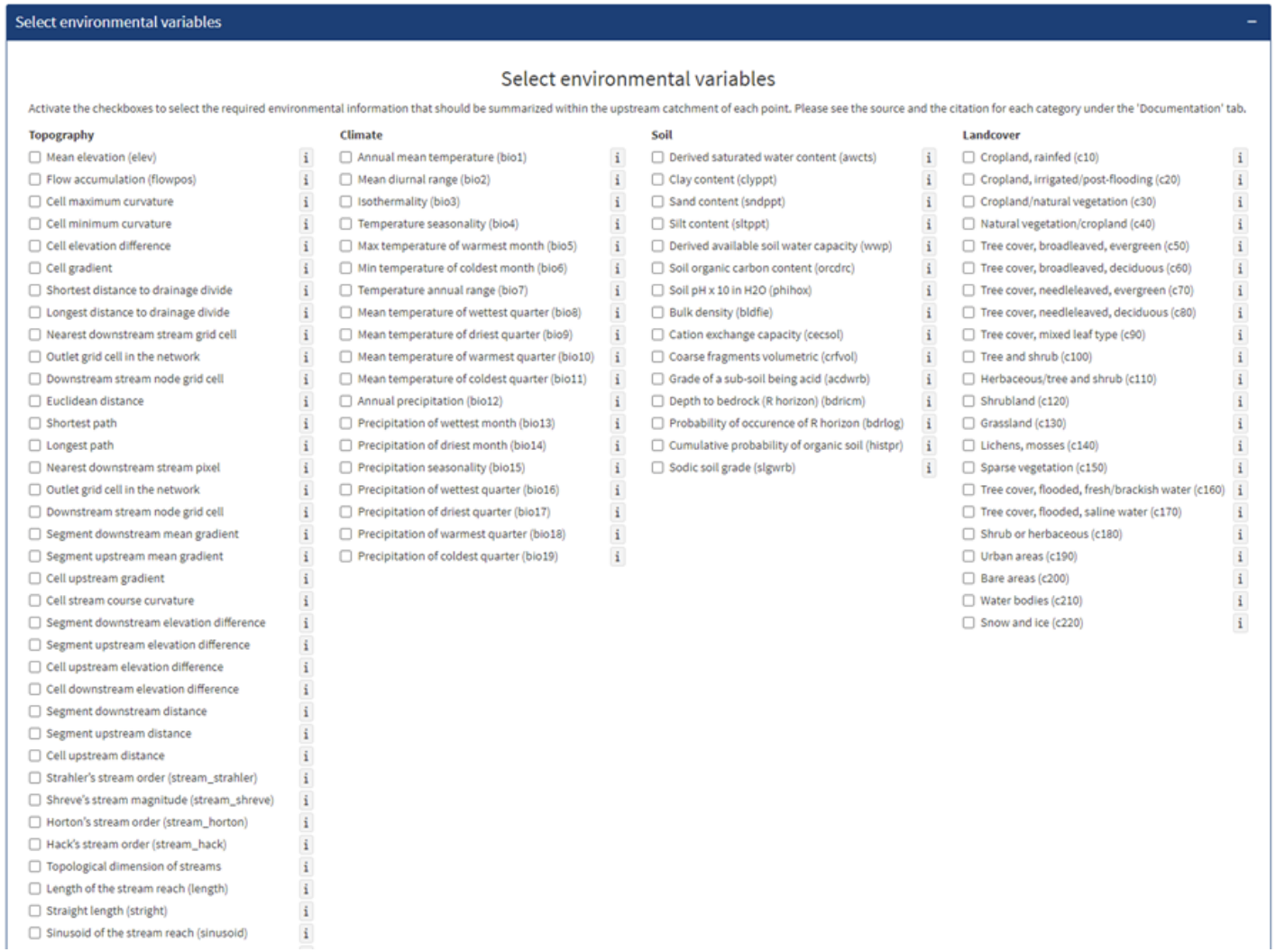
2. Snap points to stream network

3. Visualize points on the map





geofresh.org



1. Upload point coordinates

ID | lon | lat

GeoFRESH

geofresh.org

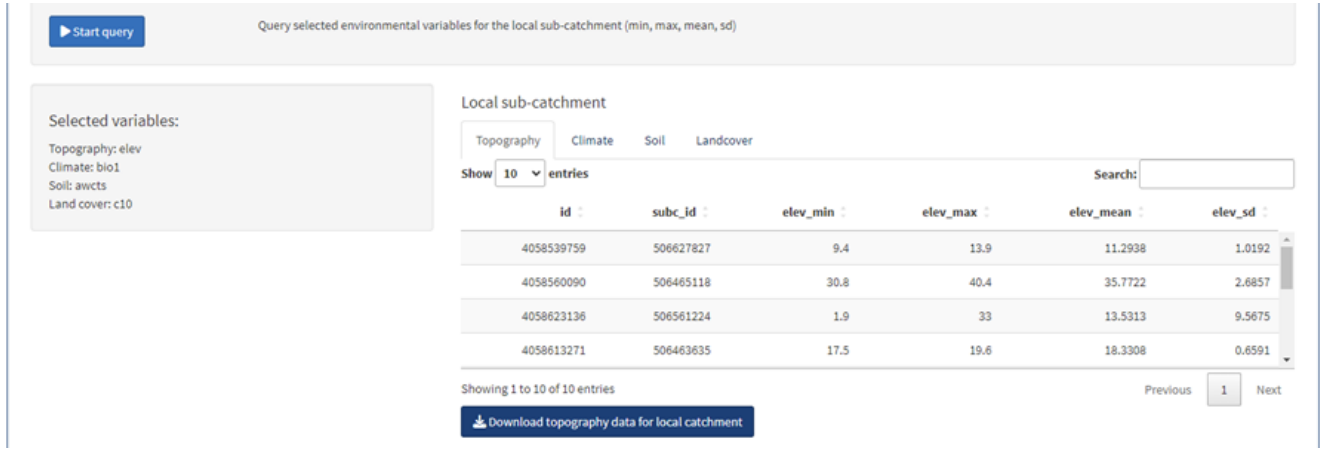
2. Snap points to stream network

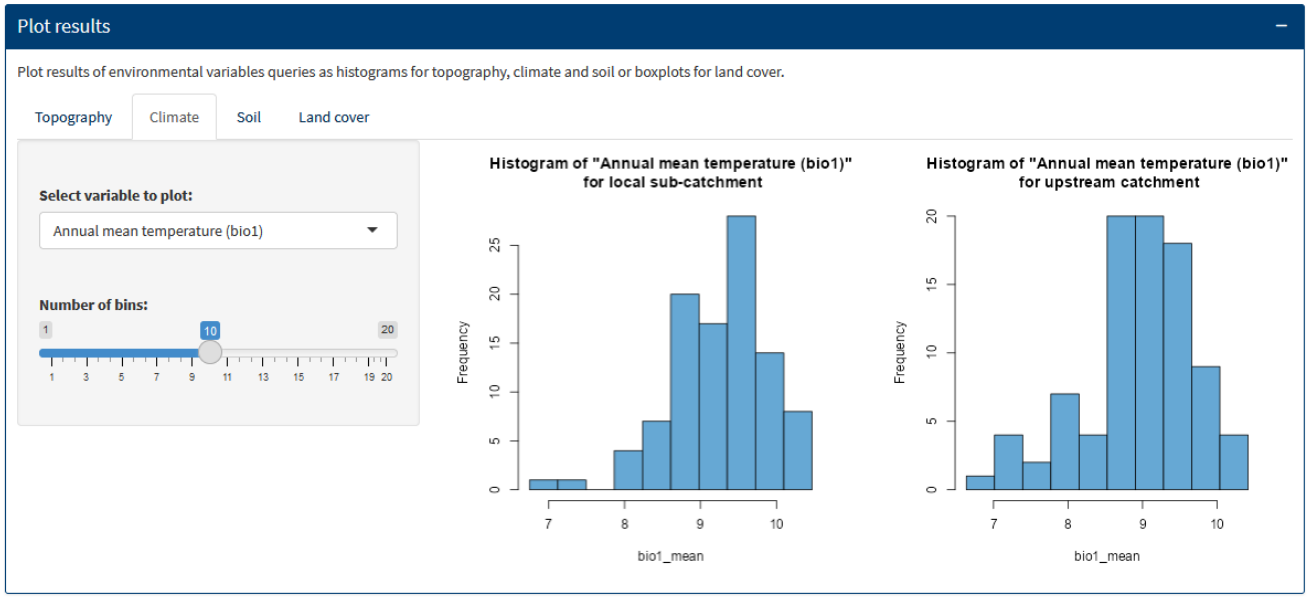
3. Visualize points on the map

4. Select environmental variables

- [ ] Topography
- [x] Climate
- [ ] Soil
- [x] Land cover

geofresh.org





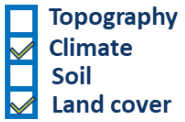
1. Upload point coordinates

geofresh.org

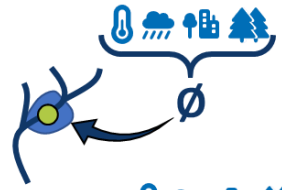
2. Snap points to stream network

3. Visualize points on the map

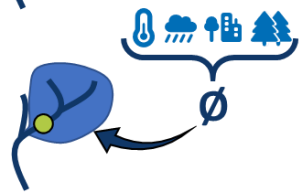
4. Select environmental variables



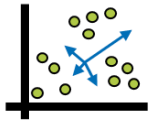
5. Query environmental variables across **local** sub-catchments



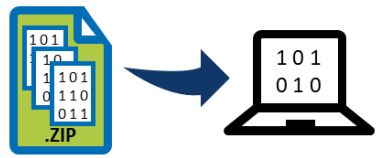
6. Query environmental variables across **upstream** catchments



7. Visualize summary statistics

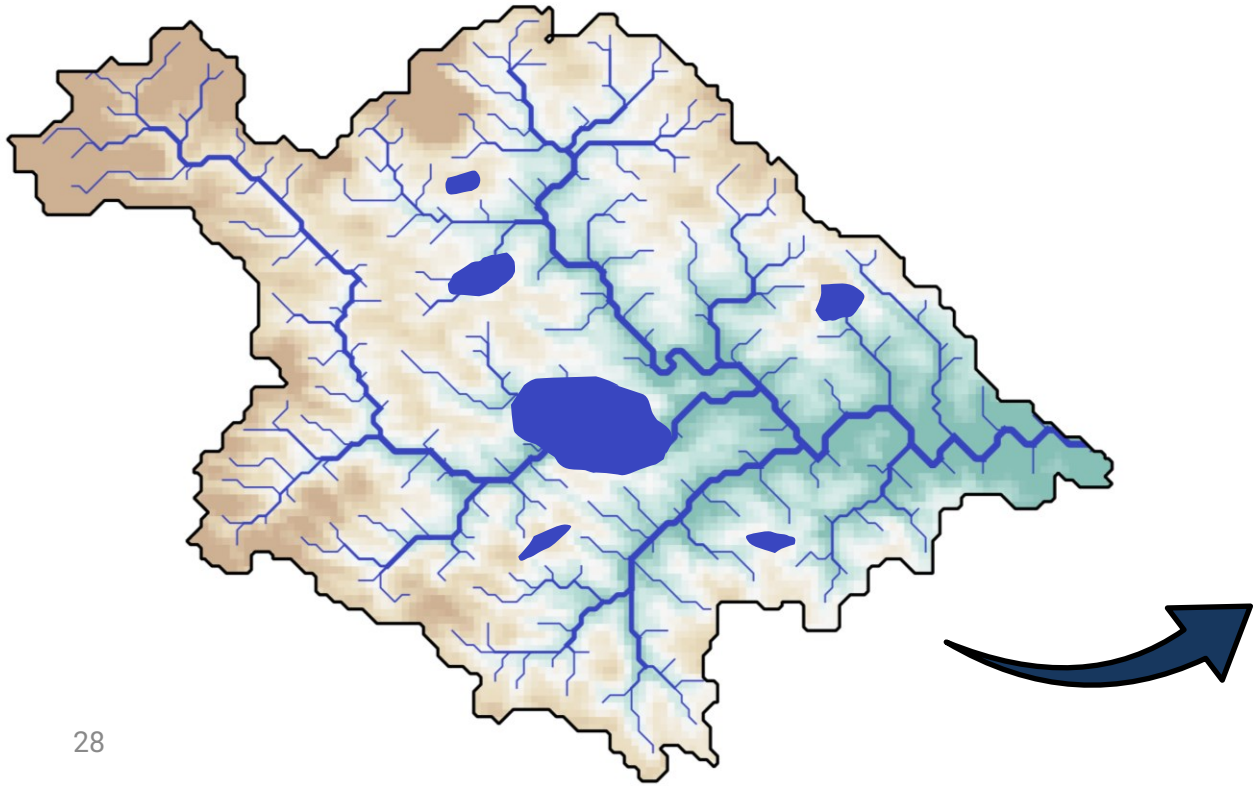


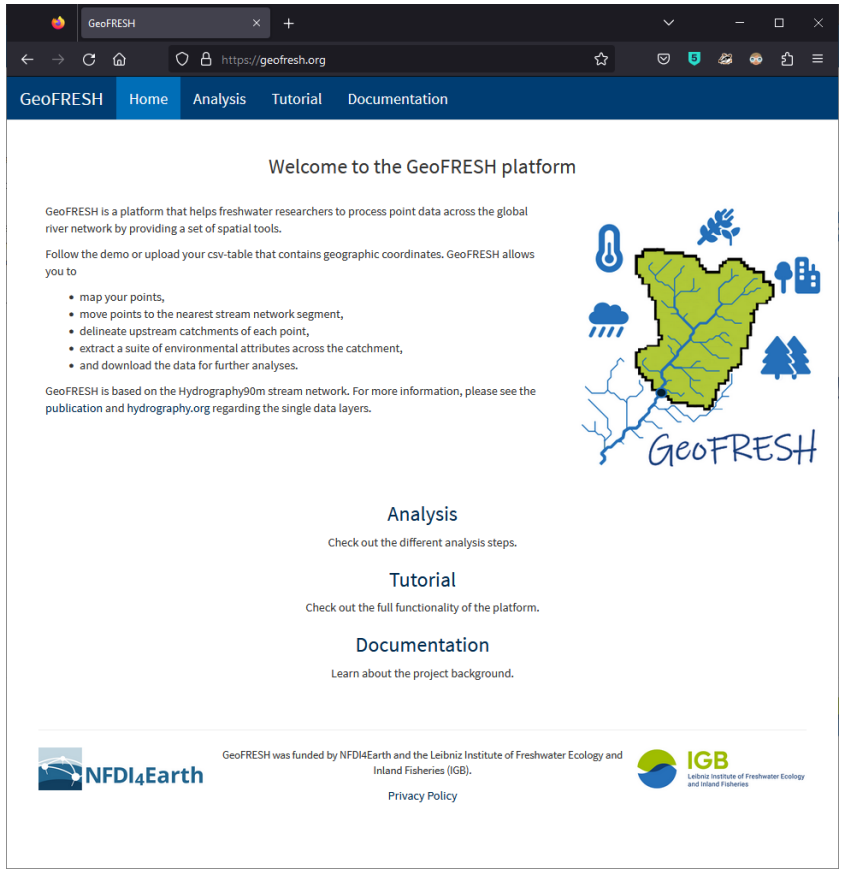
8. Download data as .csv-tables



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)





eosc | AquaINFRA

3 Virtual Research Environment | About | Logout

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

01 January 1851 *31 December 2022*

Apply temporal filter

Publication date

Thank you!

NFDI4Earth

NFDI 4 BIODIVERSITY

Leibniz Competition

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 | AquaINFRA

Funded by the European Union

Yale

SPATIAL-ECOLOGY

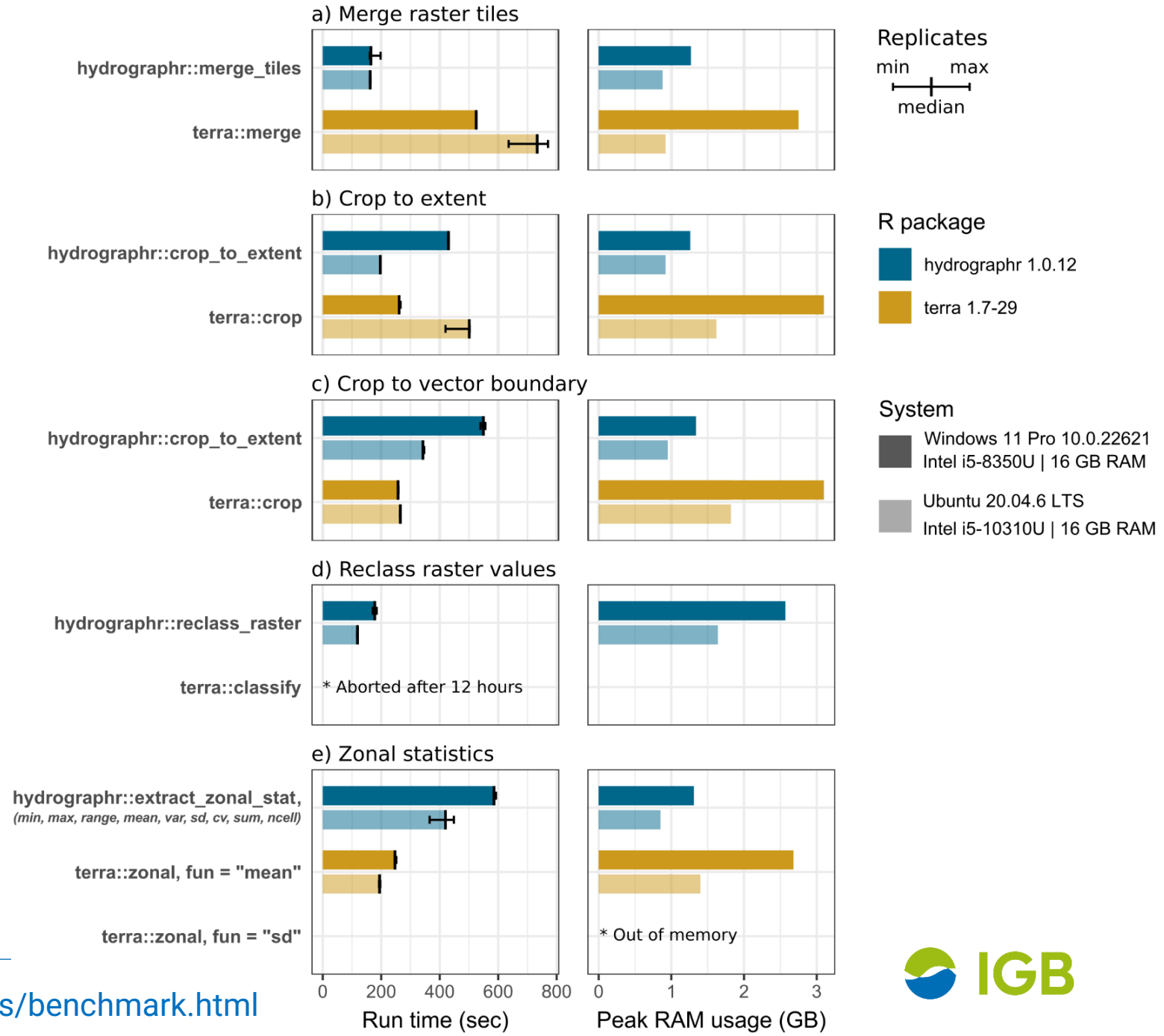
IGB

www.glowabio.org

sami.domisch@igb-berlin.de

Comparison with the terra package

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



glowabio.github.io/hydrographr/articles/benchmark.html