INTERNATIONAL JOURNAL OF DIGITAL EARTH
2024, VOL. 17, NO. 1, 2391033
https://doi.org/10.1080/17538947.2024.2391033

GeoFRESH – an online platform for freshwater geospatial data processing

Sami Domisch[a], Vanessa Bremerich[a], Merret Buurman[a], Béla Kaminke[a], Thomas Tomiczek[a], Yusdiel Torres-Cambas[a], Afroditi Grigoropoulou[a,b], Jaime R. Garcia Marquez[a], Giuseppe Amatulli[c], Hans-Peter Grossart[d,e], Mark O. Gessner[d,f], Thomas Mehner[a], Rita Adrian[a] and Luc De Meester[a,b,g]

[a]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany; [b]Department of Biology, Chemistry, Pharmacy, Institute of Biology, Freie Universität Berlin, Berlin, Germany; [c]School of the Environment, Yale University, New Haven, CT, USA; [d]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Stechlin, Germany; [e]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany; [f]Department of Ecology, Berlin Institute of Technology (TU Berlin), Berlin, Germany; [g]Department of Biology, University of Leuven, Leuven, Belgium

ABSTRACT
Freshwater ecosystems are characterized by their unique longitudinal and lateral habitat connectivity. As a result, spatial units in freshwater-specific analyses can often not be considered independent of each other. Accounting for this connectivity in modelling analyses requires advanced skills in Geographic Information Systems (GIS) for adequately processing and managing the data. To address this challenge, we developed the GeoFRESH online platform, which is available at https://geofresh.org. The platform provides a graphical, easy-to-use interface to create freshwater-specific analysis-ready data for any given location in the world, based on a high-resolution stream network (https://hydrography.org/hydrography90m/hydrography90m_layers). Users can (i) upload and visualize point coordinates, (ii) automatically assign points to the closest stream network segment, (iii) annotate the point data with a suite of 104 local and/or upstream-aggregated topographic, climatic, land-cover and soil variables, (iv) visualize summary plots, and (v) download the data in csv-format for further analyses. The platform can be expanded given its modular structure and it can serve as a key element to support freshwater science and management relying on high-resolution geospatial analyses. GeoFRESH provides a low-entry interface while being complementary to the *hydrographr* R-package, and contributes importantly to the re-usability of data as an important aspect of the FAIR principles.

ARTICLE HISTORY
Received 2 February 2024
Accepted 6 August 2024

KEYWORDS
Freshwater; connectivity; network; Hydrography90m; *hydrographr* R-package; Earth System Science

1. Introduction

Geospatial analyses of river network data processing require consideration of lateral, longitudinal and vertical habitat connectivity, including linkages of river channels with the terrestrial surroundings and groundwater (Premke et al. 2016), to account for hydrogeomorphological dynamics, sediment transport, nutrient fluxes, species migration pathways, and more, which are all important in governing ecosystem dynamics. In addition, human interventions of channel hydromorphology,

CONTACT Sami Domisch sami.domisch@igb-berlin.de

including extensive damming (Grill et al. 2019), have disrupted river networks at multiple spatial scales, which has led to massive habitat fragmentation. Spatially explicit analyses are key to describe and understand such changes, whether natural or anthropogenic, and to develop scenarios for future river ecosystem development and management (Jumani et al. 2020). A critical prerequisite for such analyses is detailed knowledge on Geographic Information System (GIS) data processing. This is true as much for analyses conducted at local scales, for example as a basis for river restoration, as it is for integrating global-scale Earth System data into river research, monitoring and management. However, the specific spatial structure of river networks and the intimate upstream-downstream and channel-floodplain linkages pose major analytical challenges. In particular, unlike spatially explicit analyses applied to land, simple data overlays of land cover, climate or dams, at a given position within a river network are insufficient to capture the potential dependencies among environmental drivers affecting river ecosystems (Schürz et al. 2023).

A possible solution to meeting this challenge would be a scalable framework that allows routing information about upstream river and catchment conditions through the hydrographical network. This spatially integrated information can then be used to assess environmental impacts, whether caused by water pollution, nutrient loading, the disruption of connectivity by dams or weirs, or flow alteration including extreme flood and drought events, at any point in the river network (Lemm et al. 2021). Such connectivity-dependent data processing frameworks for analysing river networks have been successfully used in the past (Vörösmarty et al. 2010; Jasiewicz and Metz 2011; Neteler et al. 2012). However, they confront prospective non-specialists with the need to use GIS software that can be complex, especially in the case of large datasets where elaborate data processing workflows are indispensable, requiring advanced expertise in command-line scripting. Consequently, to ensure scalability of workflows and avoiding computers to succumb to large data volumes, the choice of software and tools is critical, as is acquaintance with a particular scripting language/syntax. In addition, when the integration of data on catchment conditions is necessary (e.g. on climate or land cover, for example to calculate the fraction of forest cover within a given catchment), effective retrieval of the data from diverse online platforms might create an additional burden for users. This is because the growing number of environmental data sets are scattered and require prior assessment of data quality, availability, spatial resolution, temporal coverage and file format before being combined and analysed together with other data.

With these challenges in mind, we have started developing an online platform named GeoFRESH, which emerged from a pilot study (https://zenodo.org/records/7888389) that aimed at creating an easy-to-use and yet powerful tool that supports a broad community of freshwater scientists and managers in performing geospatial analyses of river ecosystems. Three main goals guided the implementation of the platform:

(i) Ensure easy access for all potential users, even if their technical expertise in geospatial analyses is limited.
(ii) Provide a central online hub for the actual data processing to avoid the need for massive data downloads and handling large amounts of data on the users' personal computers.
(iii) Ensure that the functionality and data-handling capacity is scalable to support workflows for both small- and large-scale analyses.

The solution we propose to meet these goals is an online platform where users can upload point data coordinates, capitalize on the latest high-resolution hydrographical data, annotate the point data with both local and upstream environmental data (i.e. add values to each data point), and download the processed data for further analyses or presentation. The resulting GeoFRESH platform provides a low-entry starting point for non-specialist users to extract and process environmental information specifically for geospatial analyses of river networks, whether focussing on local sub-catchments corresponding to the input coordinates, or targeting catchments upstream of the input coordinates, including at the global scale. The GeoFRESH platform is available at https://geofresh.org and continuously evolves by the integration of new datasets and extension of functionalities.

2. Methods & datasets

The Hydrography90m dataset (Amatulli et al. 2022, https://hydrography.org) provides the fundamental hydrographical data underlying the GeoFRESH platform. Hydrography90m is a global and seamless high-resolution river network at 90 m resolution, which is fine enough to delineate small headwater stream channels. The base data of Hydrography90m for GeoFRESH comprises 1.6 million drainage basins, 726 million unique stream segments (consisting of nodes and segments themselves) and the corresponding sub-catchments. The average size of the sub-catchments is 0.19 km^2, with a median and standard deviation of 0.14 and 0.55 km^2, respectively. Each stream segment and corresponding sub-catchment share a unique ID. Refer to Figure 1 in Amatulli et al. (2022; https://essd.copernicus.org/articles/14/4525/2022/essd-14-4525-2022-f01-high-res.pdf) for basin, segment and sub-catchment terminology.

The core idea of GeoFRESH is to provide users with local and upstream environmental information for their point records, typically sampling locations. These points are assigned to their respective stream segment in the Hydrography90m network. Users can query environmental information from four categories (topography, climate, soil and land cover) and annotate the input coordinates with a range of summary metrics (e.g. mean, range, standard deviation) of the selected variables. Finally, all data, stored in csv-files, can be downloaded in a zip-file to continue analyses on a local computer or another computational platform.

Table 1 provides a brief overview of the type of variables included in GeoFRESH. We calculated summary statistics for a total of 104 environmental variables for each single sub-catchment using the GRASS GIS software (Neteler et al. 2012). For topography, climate and soil data, we computed the mean, standard deviation, minimum, maximum and range by applying a zonal statistics algorithm based on the *r.univar* command available in GRASS GIS. For the land-cover variables, we calculated the proportion of each land-cover category in each sub-catchment.

3. PostgreSQL database and user interface

All data are stored and managed in PostgreSQL (2024), an open-source Relational Database Management System. This allows for efficient data searches and retrieval while ensuring high data

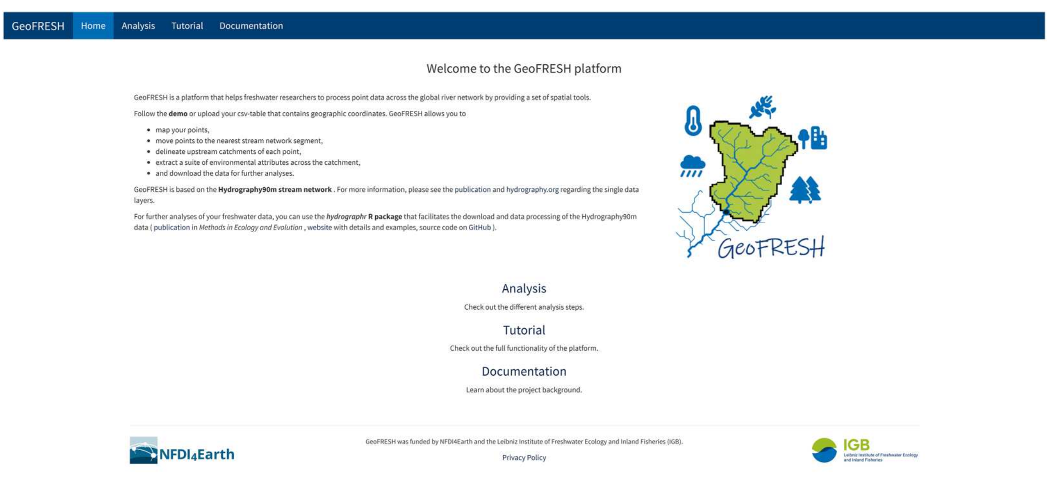

Figure 1. The GeoFRESH landing page at https://geofresh.org.

Table 1. Overview of the environmental variables currently available in the GeoFRESH online platform at https://geofresh.org. Additional variables and time-series information are currently being produced (see also the section 'Future development' below).

Description	Dataset name	Time span	Temporal resolution	Number of variables	
Global hydrography	Hydrography90m	Static	Static	48	Amatulli et al. (2022)
Climatologies at high resolution for the earth's land surface areas	CHELSA v.2.1	1980–2010	30-year average	19	Karger et al. (2017, 2018)
Global gridded soil information	SOILGRIDS	2016	Static	15	Hengl et al. (2017)
Consistent global land-cover maps	ESA Land Cover	1992–2020 (GeoFRESH: currently 2020)	Annual	22	ESA (2017)

integrity and concurrent access by multiple users. To cope with the large data volumes inherent to the description of river network structure and associated information on environmental catchment characteristics, the database tables are structured using table partitioning by regional units of the Hydrography90m dataset (see Amatulli et al. 2022 for details on regional units). Furthermore, appropriate spatial and non-spatial indexes were created. All environmental data are stored in separate partitioned tables for topography, climate, soil and land cover, while vector geometry data for stream segments, sub-catchments, basins and regional units are stored in spatial tables making use of the PostGIS extension (PostGIS 2024), which adds support for spatial data types and indexes and enables the use of spatial queries in standard SQL syntax. The *pool* package (Cheng, Borges, and Wickham 2024) is used to connect to the database by means of R code to create a connection pool.

The *dbplyr* (Wickham, Girlich, and Ruiz 2024) and *DBI* packages (Wickham and Müller 2024) are used to enable further SQL queries for uploading and processing user-defined point coordinates and for retrieving aggregated environmental data. Moreover, the *pgRouting* package (2024) serves to extend the PostgreSQL/PostGIS functionality by adding an effective geospatial database routing. These routines are used to calculate the upstream catchment for each uploaded point. To serve the data as Open Geospatial Consortium (OGC) web services (e.g. Web Map Service, WMS), database vector tables and additional raster data from the Hydropgraphy90 m dataset are registered in a GeoNode instance (https://geo.igb-berlin.de) by using GeoServer as the map server. Stream segments and sub-catchment WMS layers are integrated into the platform's mapping interface as reference layers. Users can interact with the database via an interactive web application designed to optimize user experience for simplicity and ease of use. An interface developed by using R Shiny (Chang et al. 2024) provides a framework tailored to create interactive web applications by using R as the programming language. The web application is structured into modular components, thereby facilitating updates and enabling the seamless future integration of new functionalities into the GeoFRESH platform. Altogether, this database set-up provides an effective modular structure to perform geospatial river ecosystem analyses.

4. The GeoFRESH platform

The GeoFRESH data platform is available at https://geofresh.org. It consists of four tabs that allow users to navigate to the landing, analysis, tutorial and documentation page, respectively (Figure 1). The analysis tab contains the actual workflow for data processing, whereas the tutorial and documentation tabs provide information about the functionality and background information of GeoFRESH and the underlying datasets, respectively.

5. The general workflow

The general workflow in GeoFRESH consists of eight steps guiding the user from the initial data upload to the final download of processed and annotated data (Figure 2).

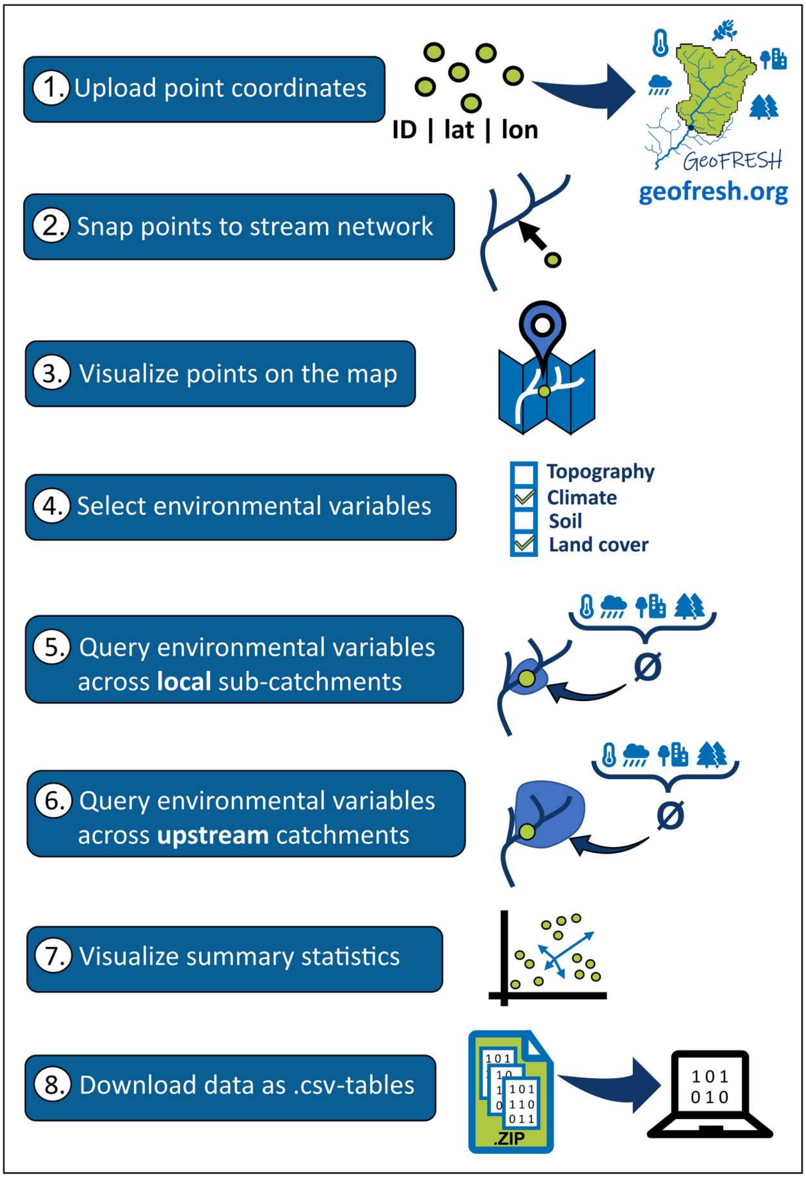

Figure 2. Schematic diagram of the GeoFRESH v.1.0 workflow, starting from the upload of coordinates, followed by the extraction of environmental data, to the final download of the processed and annotated data.

5.1. Uploading point coordinates

The point coordinate data are uploaded as a csv-table which should contain at least three columns, representing an identifier and the geographic coordinates in WGS84 (e.g. 'ID', 'latitude' and 'longitude'; Figure 3). Although the order of these columns is fixed, the column names are flexible. Alternatively, users can use the provided test dataset containing a small subset of fish occurrence records in Germany (LIB 2022), which can be loaded by a single click on the 'Load test data' button. Thereafter, users can:

- view and filter the uploaded table
- browse the points on the map, which zooms automatically to the most appropriate spatial extent

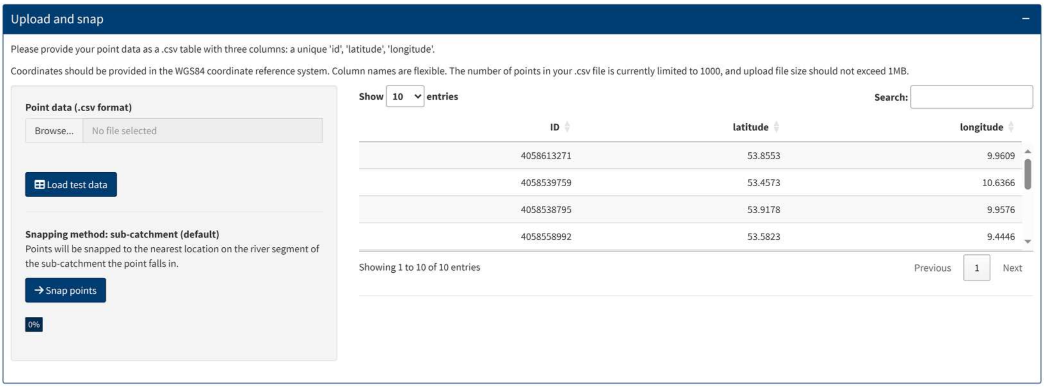

Figure 3. Screenshot of the coordinate upload section where the GeoFRESH workflow starts.

- change the background map and visualize the river segments of interest and the corresponding sub-catchments.

5.2. Snapping points to the stream network

After uploading the input data, the point records are assigned to the respective river network segment and its corresponding sub-catchment. This is a crucial step because the point locations must match the location of the stream network, which is not always the case due to inaccurate point coordinates (e.g. due to measurement errors) or the location of the modelled river network at 90 m resolution (e.g. due to the spatial resolution of the Digital Elevation Model, DEM). During the snapping procedure (Figure 4), each point record first queries the respective drainage basin ID for an approximate geographic estimation, followed by a sub-catchment ID query within the given drainage basin. The point is then snapped to the closest location on the line of the corresponding river segment. If point coordinates do not match with a stream segment, for example, if points are located in the ocean, a warning message is returned and the points are excluded from further analyses. After the point snapping procedure,

- the new latitude and longitude values (in WGS84) as well as the sub-catchment IDs are attached to the table

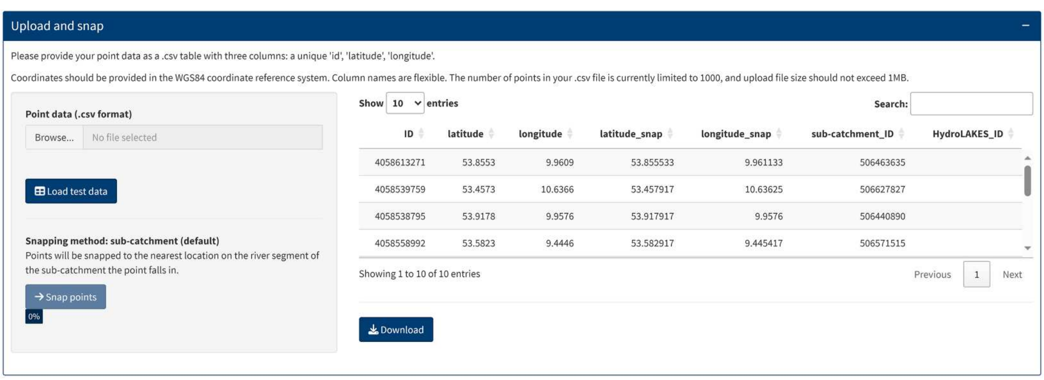

Figure 4. GeoFRESH upload and snap window showing adjusted values for latitude and longitude that have been added to the data input table after snapping point coordinates to the river network coordinates. The new table can be downloaded as a csv-file.

- users can filter the table, allowing them to perform a quick search of IDs and coordinates

5.3. Visualizing points on the map

The original and adjusted (i.e. snapped) coordinates are displayed on the map to facilitate comparisons with the original coordinates provided by the user (Figure 5(A,B)). Users can toggle between different background layers such as the stream network, the sub-catchments, or OpenStreetMap to evaluate the point location accuracy.

5.4. Selecting environmental variables

Users can select from a total of 104 environmental variables by clicking on the respective check boxes (Figure 6). Each selected variable will be considered in the subsequent data aggregation where users can select from

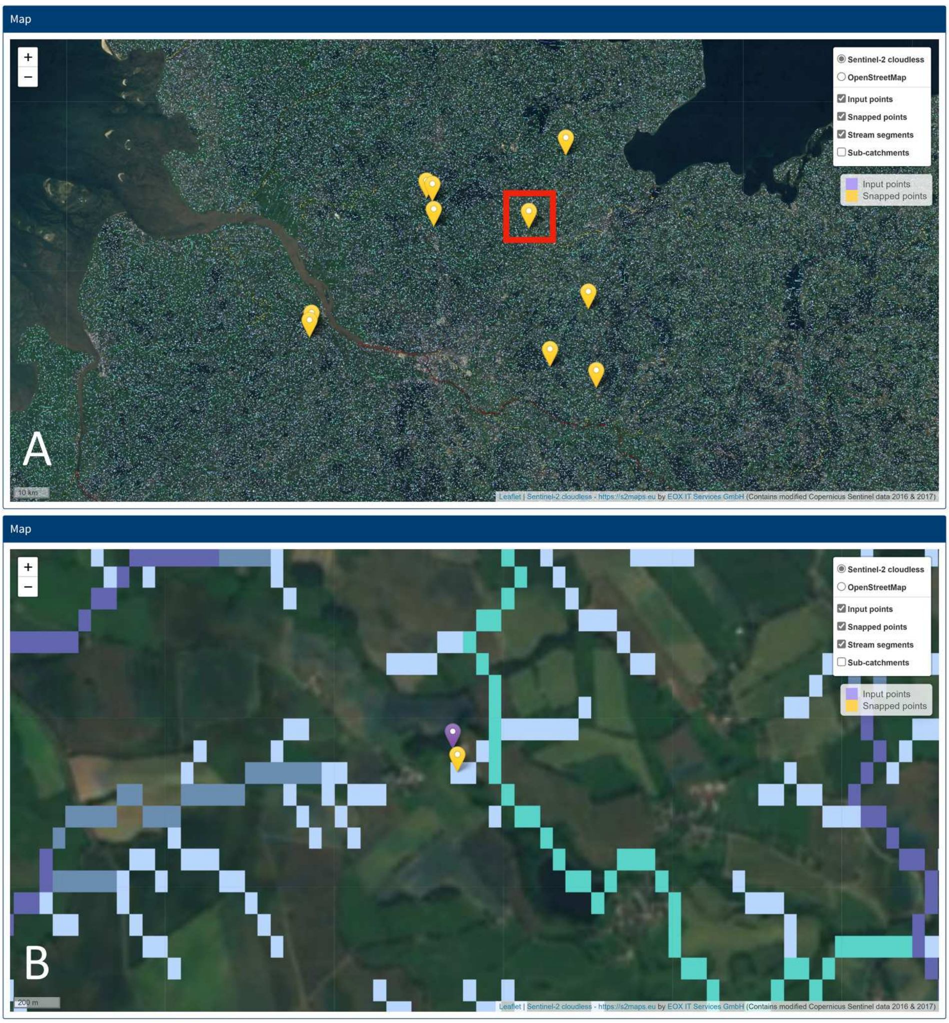

Figure 5. (A) Example of original point coordinates projected on the map. The red square represents the area shown below. (B) Adjusted point coordinates (yellow symbol) after the snapping procedure that moves the original points (purple symbol) to the closest stream network segment.

Figure 6. Options for user selections among 104 environmental variables that can be extracted for the local sub-catchment or the entire upstream sub-catchment of each point coordinate.

- 48 variables related to topography and hydrography in the Hydrography90m database (Amatulli et al. 2022).
- 19 bioclimatic variables for the period from 1981 to 2010 from the CHELSA database v2.1 (Karger et al. 2017; 2018).
- 15 soil variables from the https://soilgrids.org/ database (Hengl et al. 2017).
- 22 land cover variables for the year 2020 from the ESA Land Cover CCI Product (ESA 2017).

5.5. Querying environmental variables across the local sub-catchments

This feature retrieves the environmental information for the selected variables for those sub-catchments to which the point records are linked (Figure 7). The data is annotated to each point record individually (i.e. topography, climate, soil and land cover). In particular:

- For continuous data, the minimum (min), maximum (max), average (mean), and standard deviation (sd) values are reported for each variable (e.g. temperature) and a given sub-catchment. For categorical or ordinal data (e.g. stream order), values are reported for the respective sub-catchments.
- Users can download the annotated input tables.

5.6. Querying environmental variables across the upstream sub-catchments

This query feature enables retrieving environmental information for selected variables across the entire upstream catchment, i.e. across the upstream area of those sub-catchments to which the

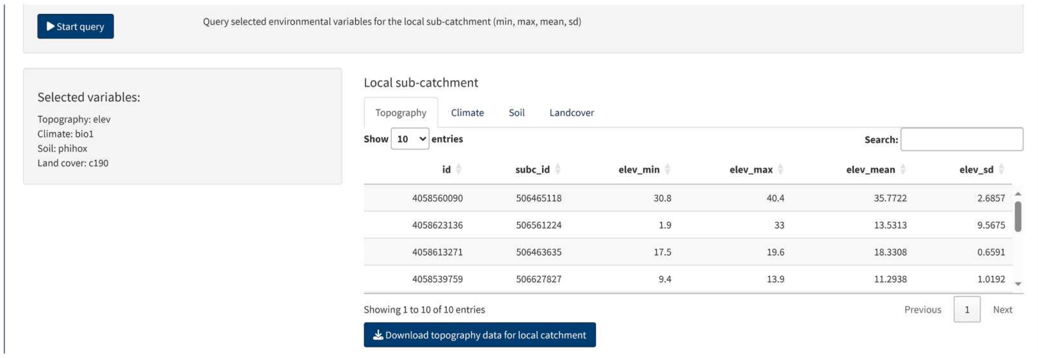

Figure 7. Query window showing the stored environmental data of the local sub-catchments after extraction of the data. The extracted data are stored in separate tables for each variable category (i.e. topography, climate, soil, land cover). The tables can be directly downloaded as csv-files.

point records were assigned (Figure 8). This functionality capitalizes on the pgRouting extension in the PostgreSQL database, using the link table of the stream segment connectivity. This means that every stream segment ‘knows’ its neighbouring segments and this allows the routing within the network. Key features are:

- Each point record on the map is individually annotated with the data of the query results.
- The average of the means for all the upstream sub-catchments in the considered area is reported.
- Categorical variables, such as stream order, are omitted from the summary statistics.
- Summary statistics are ignored for points located in the most upstream headwater segment lacking an upstream catchment. However, for the summary statistics of these upstream catchments, users can refer to the ‘local’ query.

5.7. Visualizing summary statistics

Summary statistics can be directly visualized for the selected environmental variables in pre-defined histograms and box-plots (Figure 9). Users can thus visually assess the environmental conditions at points affected by local conditions in sub-catchments or propagated downstream from the upstream catchments.

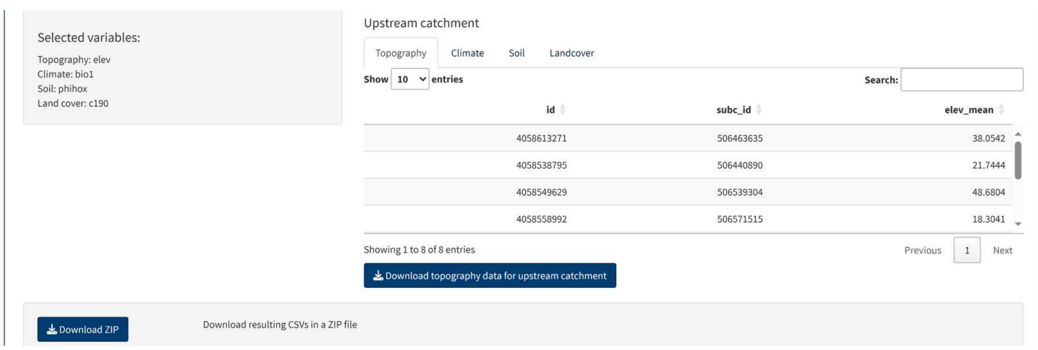

Figure 8. Example of environmental data extracted for the upstream catchments, similar to the local sub-catchments (Figure 7), and stored in separate tables for each variable category (i.e. topography, climate, soil, land cover).

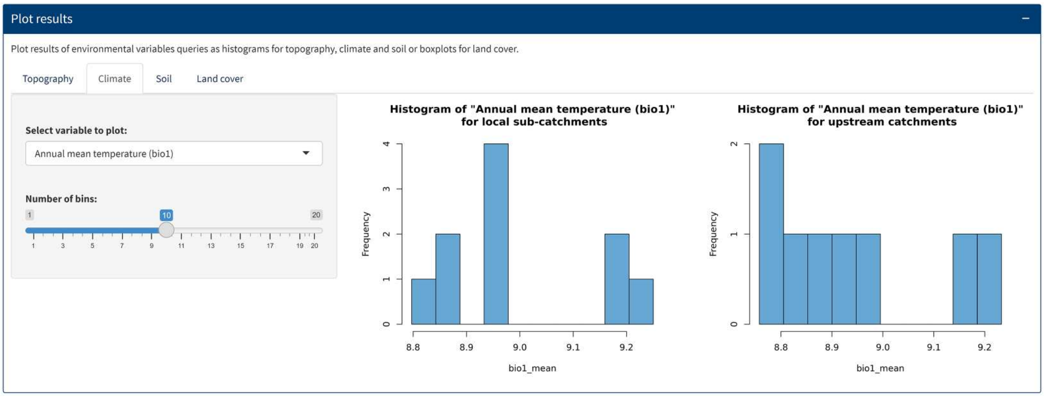

Figure 9. Direct visualization of extracted environmental data and comparisons between local sub-catchments and catchments upstream.

5.8. Downloading data as csv-files

Finally, all data can be downloaded in zip-files that contain tables as csv-files of all variable categories as well as the results of the snapping procedure. Results are kept in separate single files per category, since the csv-files contain all summary metrics. This results in a large number of columns, which would otherwise be difficult to handle in a graphical spreadsheet editor.

6. Data privacy

GeoFRESH is hosted by the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) in Berlin, Germany. Any data uploaded by a user are assigned a random ID string during processing. No cookies are employed. After users close a session, all data they provided or extracted during the session for the point records are deleted.

7. Discussion

The GeoFRESH platform is an important first step towards facilitating geospatial analyses, monitoring and modelling of river networks and associated habitats by non-specialists. It combines and arranges a wide variety of environmental data types into a single user-friendly platform, removes redundancies in the management and easy processing of (big) geospatial and environmental datasets, and minimises computational expense, thus supporting users in the scientific community and water management by providing large amounts of high-resolution spatial data on river networks. Processing of these data requires an efficient and targeted workflow, especially in view of the dendritic structure of river networks, which should best be accessible all potential users, including non-specialist and non-programmers. Given often complex GIS routines to assess connectivity and integration of environmental data in workflows, GeoFRESH greatly facilitates geospatial analyses of river networks by offering simple and yet powerful functionalities. In particular, the basic workflow of GeoFRESH presented here has high potential to ease assessments of environmental characteristics that govern material and energy flows, biodiversity and community structure, and the functionality of river ecosystems. This, together with the virtually instant data extraction enabled by GeoFRESH, as well as the ease of high-resolution analyses across large spatial scales at any given location on Earth, will provide a powerful resource serving freshwater scientists and managers, policy makers and other stakeholders in sustainable water management and monitoring of freshwater ecosystem services. For example, application of the basic GeoFRESH functionality can aid in exploring impacts of climate or land-cover change on river ecosystems, as well as in identifying drivers of water quality and availability, or species occurrences.

GeoFRESH is designed to ensure ease of use and simplicity for non-programming users, further facilitated by the open availability of all data, code and methods on Github. For geospatial analyses not currently supported by GeoFRESH, such as point-distance calculations or advanced aggregation methods, involving the use of categorical variables (e.g. the use of modes) or distance-related aggregation methods (e.g. inverse-distance weighting), the *hydrographr* R-package (Schürz et al. 2023) is available. This package allows users to apply custom data-processing workflows. Importantly, *hydrographr* uses the same Hydrography90m database as GeoFRESH, so that all basic data on river networks, including IDs, match perfectly, thus ensuring full compatibility and complementarity of *hydrographr* and GeoFRESH.

8. Future development

At present, the GeoFRESH environmental data catalogue is supplemented with additional data products that we expect to further facilitate geospatial analyses of river ecosystems across large scales, independent of particular focal regions. This includes (i) monthly discharge estimates for each stream segment, (ii) time series of climate and land-cover data over the past decades, and (iii) projections of future climate under various Representative Concentration Pathway (RCP) trajectories (van Vuuren et al. 2011). Moreover, standing water bodies are currently being integrated into the Hydrography90m dataset, which will enable integrated analyses of both riverine and lacustrine ecosystems (Messager et al. 2016). In this context, we intend to implement a separate workflow section that enables analyses of (iv) local and upstream environmental conditions potentially affecting standing waters, (v) the extent and consequences of lake connectivity along river networks, and (vi) the consideration of isolated ponds not connected to a river. Finally, we aim (vii) to integrate point connectivity analyses along river networks by capitalizing on the pgRouting (2024) extension.

The modular architecture of the GeoFRESH platform also facilitates expanding its functionalities in addition to its data catalogue. At present, we envisage to supplement the current version 1.0 with additional features such as (i) extended point-snapping functionalities, (ii) an extended interactive mapping interface for point records and the corresponding upstream catchments, and (iii) improved interoperability with the *hydrographr* R-package, particularly to enable users to tap on the online GeoFRESH functionality directly within R. Additional features can be envisaged, including requests by future users, whose feedback can be directed to the Github issues page at https://github.com/glowabio/geofresh/issues.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was primarily funded by the German Research Foundation (NFDI4Earth, DFG project no. 460036893, https://www.nfdi4earth.de). In addition, we acknowledge funding through a Leibniz Competition grant awarded to Sami Domisch (J45/2018), by the German Federal Ministry of Education and Research (BMBF grant agreement no. 033W034A), and by a grant of the Alexander von Humboldt Foundation to Yusdiel Torres-Cambas (Ref. 3.2-CUB-1212347-GF-P). Vanessa Bremerich acknowledges funding through the Leibniz Competition project “Freshwater Megafauna Futures” and Yusdiel Torres-Cambas for funding through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; CRC RESIST, SFB 1439/1 2021–426547801) and the European Union’s Horizon Europe research and innovation programme (DANUBE4all, grant agreement No. 101093985).

Data availability

All data used in the GeoFRESH platform are openly available at https://geofresh.org. See also the section ‘Methods & datasets’.

Code availability

The GeoFRESH platform is available at https://geofresh.org the source code on GitHub at https://github.com/glowabio/geofresh.

References

Amatulli, G., J. G. Marquez, T. Sethi, J. Kiesel, A. Grigoropoulou, M. M. Ublacker, L. Q. Shen, and S. Domisch. 2022. "Hydrography90m: A new High-Resolution Global Hydrographic Dataset." *Earth System Science Data* 14 (10): 4525–4550. https://doi.org/10.5194/essd-14-4525-2022.

Chang, W., J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, and B. Borges. 2024. shiny: Web Application Framework for R. R Package Version 1.8.0.9000, https://github.com/rstudio/shiny,https://shiny.posit.co/.

Cheng, J., B. Borges, and H. Wickham. 2024. pool: Object Pooling. R Package Version 1.0.2. https://CRAN.R-project.org/package=pool.

ESA. 2017. Land Cover CCI Product User Guide Version 2. Tech. Rep. Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Grill, G., B. Lehner, M. Thieme, B. Geenen, D. Tickner, F. Antonelli, S. Babu, et al. 2019. "Mapping the World's Free-Flowing Rivers." *Nature* 569 (7755): 215–221. https://doi.org/10.1038/s41586-019-1111-9.

Hengl, T., J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, et al. 2017. "SoilGrids250m: Global Gridded Soil Information Based on Machine Learning." *PLoS One* 12 (2): e0169748. https://doi.org/10.1371/journal.pone.0169748.

Jasiewicz, J., and M. Metz. 2011. "A new GRASS GIS Toolkit for Hortonian Analysis of Drainage Networks." *Computers & Geosciences* 37 (8): 1162–1173. https://doi.org/10.1016/j.cageo.2011.03.003.

Jumani, S., M. J. Deitch, D. Kaplan, E. P. Anderson, J. Krishnaswamy, V. Lecours, and M. R. Whiles. 2020. "River Fragmentation and Flow Alteration Metrics: A Review of Methods and Directions for Future Research." *Environmental Research Letters* 15 (12), 123009. https://doi.org/10.1088/1748-9326/abcb37.

Karger, D. N., O. Conrad, J. Bohner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2017. "Climatologies at High Resolution for the Earth's Land Surface Areas." *Scientific Data* 4:170122. https://doi.org/10.1038/sdata.2017.122.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2018. "Data from: Climatologies at High Resolution for the Earth's Land Surface Areas." *EnviDat*. https://doi.org/10.16904/envidat.228.v2.1.

Lemm, J. U., M. Venohr, L. Globevnik, K. Stefanidis, Y. Panagopoulos, J. van Gils, L. Posthuma, et al. 2021. "Multiple Stressors Determine River Ecological Status at the European Scale: Towards an Integrated Understanding of River Status Deterioration." *Global Change Biology* 27 (9): 1962–1975. https://doi.org/10.1111gcb.15504.

LIB. 2022. "Leibniz Institute for the Analysis of Biodiversity Change (LIB). Harmonised Freshwater Fish Occurrence and Abundance Data for 12 Federal States in Germany." *Occurrence Dataset*. https://doi.org/10.15468/c75fky. Accessed via GBIF.org on 31 January 2024.

Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. "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.

Neteler, M., M. H. Bowman, M. Landa, and M. Metz. 2012. "GRASS GIS: A Multi-Purpose Open Source GIS." *Environmental Modelling & Software* 31:124–130. https://doi.org/10.1016/j.envsoft.2011.11.014.

pgRouting. 2024. pgRouting Contributors: pgRouting of PostGIS and PostgreSQL Geospatial Database, Version 3.6.1, https://pgrouting.org/ (last access: 22 May 2024).

PostGIS. 2024. PostGIS Development Team: PostGIS Spatial Extension for the PostgreSQL Relational Database, Version 3.4.1, https://postgis.net/ (last access: 22 May 2024).

PostgreSQL. 2024. The PostgreSQL Global Development Group: PostgreSQL Relational Database System, Version 16.1, https://www.postgresql.org/ (last access: 22 May 2024).

Premke, K., K. Attermeyer, J. Augustin, A. Cabezas, P. Casper, D. Deumlich, J. Gelbrecht, et al. 2016. "The Importance of Landscape Diversity for Carbon Fluxes at the Landscape Level: Small-Scale Heterogeneity Matters." *WIRES Water* 3 (4): 601–617. https://doi.org/10.1002/wat2.1147.

Schürz, M., A. Grigoropoulou, J. García Márquez, Y. Torres-Cambas, T. Tomiczek, M. Floury, V. Bremerich, et al. 2023. "hydrographr: An R Package for Scalable Hydrographic Data Processing." *Methods in Ecology and Evolution* 14 (12): 2953–2963. https://doi.org/10.1111/2041-210X.14226.

van Vuuren, D. P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, et al. 2011. "The Representative Concentration Pathways: An Overview." *Climatic Change* 109 (1-2): 5–31. https://doi.org/10.1007/s10584-011-0148-z.

Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, et al. 2010. "Global Threats to Human Water Security and River Biodiversity." *Nature* 467 (7315): 555–561. https://doi.org/10.1038/nature09440.

Wickham, H., M. Girlich, and E. Ruiz. 2024. dbplyr: A 'dplyr' Back End for Databases. R package version 2.4.0, https://github.com/tidyverse/dbplyr. https://dbplyr.tidyverse.org/.

R Species Interest Group on Databases, Wickham, H., and K. Müller. 2024. DBI: R Database Interface. R Package Version 1.2.1. https://CRAN.R-project.org/package=DBI.