

Review of the manuscript:

“Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution”

Short summary

In the data paper “Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution”, the authors present an aggregation of available global environmental datasets resampled for the sub-catchments of the Hydrography90m dataset. The variables include topography, hydrography, present and future climate variables, land cover, soil variables, aridity, and modelled streamflow.

In the **introduction**, the authors describe the need to protect freshwater biodiversity and habitats. To conduct globally standardised analysis and modelling studies, scientists require environmental data with a very high spatial resolution and global extent. The Environment90m dataset, based on the Hydrography90m stream network (Amatulli et al., 2022), should fill this gap.

In the **Environmental Data** section, each of the individual underlying datasets (all of which were published previously) is presented. An overview table is given for each dataset: Stream network data; climate; land cover; soil; elevation; stream flow; global aridity; and potential evapotranspiration.

In the **Calculations** section, the authors describe the dataset subsampling procedure. They also describe how the data is accessible to users.

The **Case study workflow** provides an example of how to use the dataset to predict the distribution of fish species in the Danube region.

The article ends with a **conclusion** that summarises applications of the dataset and potential possibilities to extract also data for lotic habitats.

Main concerns

After carefully reading the research article “Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution”, I see an immense value in the presented dataset. The data presented is not new itself, but the consistent aggregation and re-sampling on a global scale is of high value for further studies. Therefore, it will be of high interest to a huge number of users. The dataset is very well accessible by tools provided to access the data via an R package (hydrographr), an online platform (GeoFresh), or direct downloads seem very helpful and are well documented with vignettes.

Review - "Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution"

However, I have some suggestion about how to improve the structure und presentation of the dataset, which I will outline below:

Structure & Presentation of the dataset

The manuscript would benefit from a clearer and more consistent structure.

To improve the structure, I would suggest the following new sections:

- Consider creating a dedicated section titled *"Accessing the Data"*. Currently, information about data access is scattered across the Introduction, Calculation, and Case Study sections. Consolidating this content would improve readability and help users locate key information more easily. This section could also include a more detailed explanation of the custom function in the hydrography R package.
- Consider adding a separate section for *Applications* rather than embedding these points within the Conclusions. This would allow for a more focused discussion and better highlight the broader relevance of the dataset and tools.
- A more in-depth discussion of the dataset would strengthen the manuscript. For example:
 - What are the implications of the sub-sampling effort?
 - Could the dataset be expanded in future versions?
 - The absence of water chemistry parameters (e.g., nutrients) is notable, as these are key drivers of freshwater biodiversity. A brief discussion of this limitation and its potential impact on certain taxa would be valuable.

To improve the understandability of both the methods and the dataset, I suggest reordering the relevant sections: Start with the Calculations and then present the Environmental Data including the base datasets, selection criteria, pre-processing steps, and the final derived variables.

- Reason: Currently, the presentation of the dataset and the methods is interwoven, which leads to confusion. The original input datasets are introduced first, but the final dataset - including all derived summary statistics - is not clearly distinguished. This makes it difficult to grasp what is actually included in the published Environment90m dataset.
- Suggestion: I suggest reordering the section to first explain the resampling and calculation procedures (e.g., how zonal statistics, proportions, or other metrics were computed for the sub-catchments), and then introduce the Environment90m dataset (and it input data) as the result of these operations. This would help readers better understand the transformation from raw input data to processed variables. I also recommend updating the tables to include the available summary statistics for each variable. This would make the tables more informative and clearly reflect the contents of the final dataset.

Please check that all information is given within the respective section:

- For instance, the section titled *"Case Study Workflow"* begins with a discussion of the R package's usefulness, rather than introducing the case study itself.

Language and readability

There is potential to improve the readability of the manuscript. Some sentences are overly long and complex, which makes the text difficult to follow (see examples in the specific comments). Shorter sentences and a more active voice would enhance clarity and accessibility. The language check should be done also beyond the specifically mentioned sections in the specific comments.

The in-text citation style is inconsistent. Some references include brackets around the year, while others do not. In many cases, brackets around the entire citation are missing, which disrupts the flow of reading. Please ensure that the citation style is consistent throughout the manuscript and adheres to the journal's guidelines.

Figures and tables are often hard to read due to small font sizes, unclear terminology, or lack of explanatory notes (see specific comments).

Specific comments

Title

Two thoughts about the title of the paper:

- The current title emphasizes "spatial freshwater biodiversity science". I would encourage the authors to consider whether the dataset might also be relevant to other fields, such as hydrology, landscape ecology, or environmental modelling. If so, broadening the scope of the title could help reach a wider audience and better reflect the dataset's potential applications.
- The word "*spatial*" appears twice in the title, which may be redundant. Removing the first occurrence could help streamline the title without losing clarity.

Introduction

L19: "Freshwater biodiversity is among the terrestrial and marine realms most at risk." -> I do not properly understand the meaning of the sentence. It compares biodiversity (in freshwater) with realms (terrestrial and marine)? Please re-formulate more clearly. (e.g. "Freshwater biodiversity is considered to be more threatened than biodiversity in terrestrial and marine ecosystems.")

L19: Brackets around the citation are missing.

L19-23: The sentence is difficult to understand due to its length and complexity. I recommend streamlining and shortening it to improve readability. Breaking it into two or more sentences could help clarify the intended message.

L23-26: I suggest separating the question from the rest of the sentence with ":" – e.g. "(...) are required to answer the question: which areas should be prioritised for protection?"

L39-43: This is also a very long sentence. I can hardly understand it because of its length. Please streamline and shorten.

Review - "Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution"

L45: "would lump the data" -> is this a standard term or colloquial language?

L50-51: The sentence "*Sub-catchments consist therefore of the natural units in freshwater ecosystems and allow encompassing also riparian areas and aquatic-terrestrial linkages.*" is difficult to interpret. The term "*natural units*" is unclear — could the authors clarify what is meant by this? Does it refer to ecological boundaries, hydrological divisions, or something else?

Additionally, the word "*therefore*" is typically set off by commas when used in this context. A revision for clarity and grammatical correctness is recommended.

L57: Why is here a question mark?

L81: The following mentioned link to the exemplary lake vignette is not working: https://glowabio.github.io/hydrographr/articles/case_study_lake_workflow.html.

Environmental Data

L 83-85: When I was reading the paper for the first time I was confused about the content of the section, whether this section refers only to the input data available from Hydrography90 and other datasets or also includes the processed variables used in this paper. Therefore, I suggest changing the following sentence "The following describes the underlying environmental data to derive the Environment90m dataset." into "The following section describes the foundational environmental datasets used to generate the Environment90m dataset."

General comment on spatial resolution: I struggled to understand the spatial resolution descriptions throughout the manuscript. For example, if you refer to a resolution of "90 m" (L. 87), while in table 1 it is written as "90 m²". This creates confusion: If 90 m refers to the length of one pixel, then the actual area covered by a pixel would be 90 m × 90 m = 8100 m². This same issue appears with other datasets as well (e.g. land cover, soil). Please clarify whether the resolution values refer to pixel length or pixel area and ensure consistency across the manuscript and tables.

L87 ff: Be consistent in formatting the data sets in italics.

L91-92: Could you briefly explain how the 726 million sub-catchments can be identified and located? It would be helpful to know where users can find their IDs, locations, and areas, especially for working with the dataset.

L99-100: The sentence suggests that a "combination" of three SSPs and three GCMs was used to derive the climate variables. Could the authors clarify what this means in practice? Was the mean calculated across all combinations of SSPs and GCMs, resulting in a single averaged output? Or were individual outputs generated for each SSP-GCM pairing? Please describe it more precise. And is this a step done by the authors (and should be described in the Calculations) or is this an aggregation of the original dataset?

L 103-105: Please clarify whether the data refers to land use, land cover, or both. The terminology should be consistent throughout the manuscript to avoid confusion. It seems that a reference to the

Review - “Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution”

data source is missing in this section. Including a citation or link to the original dataset would improve transparency and reproducibility. Additionally, it would be helpful to explain which categories were aggregated and how.

L110: The sentence mentions weighting based on soil depth, but it’s not entirely clear how this was implemented. Was the variable weighted by the depth of each respective soil layer?

L118: Why are annual stream flow data not provided, along with mean, maximum, and minimum annual values? This type of information could be highly valuable for environmental modelling, especially when assessing long-term hydrological patterns or linking stream flow to land use changes. Since land use data is provided annually, aligning the temporal resolution of stream flow data would improve consistency and facilitate integrated analyses. I recommend including these annual metrics or explaining why they were omitted.

Calculations

L130: Please remove the comma.

L 140: Remove the period within the brackets.

L139-140: Could the authors clarify why interpolation was not applied? In many cases, interpolation can help fill gaps or smooth spatial data. A brief explanation of this decision would help readers understand the methodological choices and any limitations that result.

Case study workflow

The current presentation of the case study reads more like a methods tutorial than a full use case. If the intention is to present a complete use case, I suggest including:

- A short introduction to the ecological or scientific relevance.
- A summary of results.
- A brief discussion of those results.

If, however, the goal is to present only the workflow, then the link to the provided vignette is maybe sufficient, as it allows for code visibility and intermediate outputs.

Tables

The descriptions (column description) within the individual tables were not always understandable, some explanations of abbreviations of terms like “focal grid cell” or “Scale” are missing. See individual comments below.

Table 1

- Spatial resolution: Is the resolution really 90 m²? Based on my understanding, each pixel in Hydrography90m has a length of 90 meters, which would result in an area of 90 m × 90 m = 8100 m². Please clarify this.

Review - “Environment90m - globally standardized environmental variables for spatial freshwater biodiversity science at high spatial resolution”

- Scale = 10^6 : The meaning of this scale value is unclear. Does it refer to map scale, data normalization, or something else? A brief explanation would be helpful.
- Focal grid cell: The term is not clearly defined. I had assumed the data is aggregated at the sub-catchment level rather than at individual grid cells. Please clarify what is meant by “focal grid cell” in this context.

Table 2

- The meaning of Scale and Offset (e.g., “Scale = 0.1, Offset = -273.15”) is not explained. Are these used for data transformation or unit conversion? If so, please provide a short note or example.
- “annual precipitation”: Consider capitalizing the first letter for consistency with other entries.

Table 3

- What is the meaning of the numbers in the description?
- “Water bodies”: Is there a possibility to differentiate between types of water bodies (e.g., lakes, rivers, streams)? This could add valuable ecological context.

Table 4

- The description is difficult to understand due to unexplained abbreviations. Please avoid using abbreviations unless they are defined in the table caption or footnotes.

Table 5

- temporal resolution and time range are missing

Figures

Figure 1

This is a visually engaging and informative overview figure. However, I recommend increasing its size for better readability (maybe some re-arrangements of the elements is needed). Currently, the font size is quite small, which makes it difficult to interpret some of the details.

Figure 2

I do not see that this figure adds new information beyond what is already presented in Figure 1.

Figure 3

This figure is also difficult to read due to its small size and dense content. Enlarging the figure and improving the layout or font size would help make it more accessible to readers.

Supplementary material

I appreciate the inclusion of multiple vignettes, tutorials, and external web resources linked in the paper. These additions significantly enhance the usability and accessibility of the dataset and tools, and they provide valuable guidance for potential users. Well done!