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

The manuscript presents Environment90m, a valuable new dataset for global freshwater research. By providing globally standardized environmental variables at high spatial resolution, this dataset addresses a key limitation in large-scale studies of freshwater ecosystems - the lack of consistent, high-resolution environmental data. The integration with the hydrographr R package is particularly commendable, as it facilitates data access and analysis at large scales.

However, before recommendation, I have several concerns regarding the clarity, structure, and presentation of the paper.

Thank you for the constructive comments which we took all into account while revising the manuscript.

Major Concerns

Manuscript structure: The structure of some sections should be revised for better logical flow (see detailed comments below).

We have done a general re-structuring of the manuscript. With this we have improved the readability and general flow of the manuscript.

Readability: Language and grammar require improvement. Sentences are frequently too long and difficult to follow. A language edit is recommended.

We checked the manuscript with special attention to grammar and improvement of the sentences to make them shorter, clearer and easy to follow.

hydrographr package updates: The extent of modifications and extensions made to the hydrographr R package is unclear. A concise summary of newly added functions - beyond those demonstrated in the case study - would be highly valuable to researchers.

We added in the “Accessing the Environment90m dataset” section (2.4) a table with a description of the new functions now available in the hydrographr R-package.

Minor Concerns

Citations often appear incorrectly formatted or are replaced by “?”, suggesting broken links to the bibliography. These should be checked and corrected.

All references have been checked and corrected. The issue was due to Laxex-libraries which did not insert (the correct) citations in the pdf document.

The vignette link is not functional. I suggest including “Environment90m” in the vignette title (e.g., “Case study - Danube Basin (Environment90m)”) to make it more easily discoverable.

Now the link to the vignette is working fine. We have renamed the title to the suggested one by the reviewer “Case study - Danube Basin (Environment90m)”. We agree that it increases the chance to be discovered and used.

In the case study (vignette), the paths (working directory setup) appear inconsistent and worked only after adjustments; also, “flow” may need to be replaced by “accumulation” in the function:

Thank you for spotting this, there was a trailing slash “/” missing when setting up the working directory path. This has been fixed now.

download_hydrography90m_tables(subset = c("flow --> accumulation?", "length",
"slope_grad_dw_cel"),
…)

The abbreviation of “flow” has been changed to “accumulation”.

Section-Specific Comments

Introduction

Line 23 - 30: This paragraph is hard to read and it seems establishing a baseline is a major motivation to assemble this data set. I suggest to elaborate on this, as it is not quite clear to me what this baseline is referring to.
We split the sentences and checked the legibility again. We also added more text on the baseline, which now reads as follows: “Addressing this question requires at minimum a detailed baseline of the present-day spatial distribution of the environmental characteristics of freshwater habitats. Only after establishing a baseline, the environmental changes, or changes in biodiversity can be measured and quantified.”
We hope that this resolves the question, and we are happy to add more text if needed.

Paragraph 2 (Lines 31 - 54): This section currently focuses on methodological difficulties rather than the broader relevance of the dataset. Consider moving this discussion to the Calculation section. Instead, emphasize the scientific and practical value of integrating stream networks with climatological, land cover, and soil data. I.e., I suggest to focus less on the technical challenges and more on the possibilities once these challenges are overcome.

Thank you for this comment. This section reflects our experiences in having to convince others why the choice of spatial units and the spatial resolution matters in spatial freshwater biodiversity science (it may seem more than obvious to others). We elaborate on the network continuum and explain how the environmental data has to be attributed to the respective stream

segments and sub-catchments, but it is true that we do not explain the rationale of this for each variable category. If this is needed, we are happy to add more text. Currently we explain the advantages in the Discussion and if needed.

Paragraph 3 (Lines 55 - 60): The comparison with existing datasets is useful but remains vague. Clarify how Environment90m advances beyond these products and articulate the specific knowledge gains it enables.

We added a paragraph to highlight the advantages of Environment90m over the existing datasets. Environment90m takes advantage of the added values provided by the Hydrography90m dataset. In particular, the detailed delimitation of upstream sub-catchments and therefore the availability of environmental data for these areas, normally absent in global and regional studies.

Line 81: The link is not working.

The link has been updated and is now functional (https://glowabio.github.io/hydrographr/articles/case_study_lakes.html).

Environmental Data

Section 2.1: The stream network data are already described in the Hydrography90m publication and may not need to be reiterated here. A concise reference to that paper may suffice.

We suggest to leave the description and table of the Hydrography90m dataset as it is right now, mainly to have consistency between this dataset and the other datasets. Also, it would facilitate users of the dataset or the R package to have one unique reference to find the description, and standard abbreviations of the underlying variables and datasets used. Finally, this paper can be considered stand-alone by including a brief description of the Hydrography90m data, which forms the backbone of the Environment90m dataset.

Section 2.2: Elaborate on the use of these 3 GCMs and why a combination of three models was Used.

We made a selection of 3 global circulation models (GCMs), where we selected 3 SSPs and two time periods. The rationale behind this selection is that the pair-wise combination of GCMs-SSPs for the two time periods covers a wide spectrum of short and long term future projections, considering different impacts given social, technological, economical and environmental changes. We believe that this diversity of options allows capturing and evaluating the uncertainties in studies that employ climate projections. We included this explanation when introducing the CHELSA dataset. In the conclusions we also mention future developments of the database where we expect to use more GCMs and also include the 2011-2040 period as well.

Section 2.3: How were the 22 categories selected from the original 37 ESA categories? Does the land-use data have a temporal resolution (for the years 1992 to 2020)? Please be consistent when referring to land-use or land-cover data.

The 22 categories were selected based on their consistency in global coverage over the entire time period outlined in the Land Cover CCI product user guide (CCI, 2017).
CCI, E. L. C.: Product user guide version 2.0, UCL-Geomatics: London, UK, 685, 2017.

The original categories of this dataset have been created in a hierarchical manner with different levels. Level 2 is meant to be used at regional scales and level 1 at the global scale. We adopted level 1 categories but added the following explanation in the caption of the table to clarify how the categories level 2 were aggregated: "...the numbers within parenthesis in the description make reference to the coding of the categories at level 2".

In the caption of the land cover table we explain that the numbers within the parenthesis in the description of each category are the codes taken from the original land cover categories used to aggregate the final categories. Also, in the description of the dataset, the temporal resolution is specified, as annual resolution, and therefore the data can be retrieved for each year within the 1992-2020 period. We also fixed the term to land cover data throughout the manuscript.

Section 2.4: Why did you decide to integrate over all available soil depths?

The original data is provided in different soil depths. At first we needed to decide if we prepare the data per soil depth or calculate the average. The calculation per soil depth would have been very computationally intense and selecting only one depth would generate some potential bias for each of the soil variables. Therefore, we decided to calculate the weighted average for each soil property following the GlobalSoilMap specifications as described in Arrouays et al. (2014).

Section 2.6: I suggest to also provide the variance of stream flow over the selected time period.

We agree that providing the variance of the selected time period and also the yearly estimates would be very beneficial for freshwater research. Given that the time-frame for calculations is long, we will leave these variables out for this version of the Environment90m dataset, but we will include them, among other new variables (see at the end of the discussion section), once the new calculations are finished.

Section 2.7: How exactly was AI and PET modeled? Please elaborate on the process.

We are aware that the underlying datasets and how they were created is important for potential users. At the same time, the descriptions of how the CHELSA climate, soilgrids, FLO1K and AI/PET were created would be beyond the scope of our manuscript. We opted for short descriptions in the table which briefly describes how the variables were calculated. We processed AI and PET using the resampling procedure explained in section 2.1 to calculate the summary statistics for these variables.

Calculations

Include a short discussion on how the varying spatial resolutions of the underlying datasets affect Environment90m applications and interpretation - particularly in small headwater catchments.

This discussion is very important and we have added text in the Discussion section to develop these ideas. We discussed the advantages of aggregating the underlying datasets (i.e. calculating the summary statistics) in comparison to using the original values, in terms of variable standardization and saving computing times for regional and global analysis. This directly also took us to discuss the advantages of using the sub-catchments instead of cell grids as units of analysis given the ecological significance of the sub-catchments for freshwater biodiversity. We finally highlighted the accessibility of high resolution environmental data available in Environment90m for headwater sub-catchments, which are normally neglected or aggregated to larger basins, and open a window of research to investigate the relevance of these areas for freshwater conservation.

Use a consistent notation for spatial resolution (either 90 m or 90 m²).

We have changed all resolution notation to e.g. 90m.

Case Study Workflow

The vignette link is not working.
We have corrected the link to the vignette https://glowabio.github.io/hydrographr/articles/case_study_lakes.html.

The case study is an excellent addition. However, please include a (short) dedicated section summarizing the new functions added to hydrographr for handling Environment90m data, possibly including a summary table.

We decided, based also on the suggestions from Reviewer 1, to include the description of the new functions of the hydrographr R-package, in the section: "Accessing the Environment90m dataset". We also included a summary table of the functions (Table 8) following the same design as in the hydrographr package manuscript.

Conclusion

It is not always clear whether the studies cited used Hydrography90m or Environment90m. Please clarify.

We have now a dedicated section (i.e., section: “Applications”) to showcase the studies where the Environment90m dataset has already been used. We clarified the narratives by explaining the data used as inputs for this case studies are coming from the Environment90m datasets and that in some cases they retrieved from the large tables some of the variables originally available in the Hydrography90m dataset, as referred in Table 1. We wrote this more clearly in the text.

New functions for lake processing are introduced only here; these should be documented earlier in a dedicated section.

We added to the manuscript a new dedicated section to illustrate a second case study related to lake processing (i.e., section: Case study: environmental characterization of lakes’ upstream basins). In this workflow we describe the new functions available in the hydrographr Rpackage for lake processing and show how to use Environment90m to environmentally characterize the upstream basins of the stream network and lake intersections and the upstream basin of the lake outlet.

Figures and Tables

Tables 1–7: Ensure uniform font size.
We have changed the fonts of all tables to have the same size.

Figure 1 & 3: Captions should be more descriptive and self-explanatory.
The caption of these two figures have been improved by expanding the descriptions of the workflows.

Overall Assessment

Environment90m represents an important contribution to global freshwater biodiversity science. With clearer presentation, improved language, and clearer documentation of the newly added hydrographr functions, this dataset will likely become a foundational resource for future large-scale aquatic research.
Thank you once again for the very positive and constructive comments, which helped us to improve the manuscript.

Citation: https://doi.org/10.5194/essd-2025-399-RC2