

## Response to Reviews

*We thank the reviewers for their thorough and constructive evaluation and feedback. We have tried to respond to all comments as well as possible. Please find our reply to RC2 below.*

### RC2

This is my first review of the manuscript by Färber et al. The manuscript addresses a relevant topic and aligns well with the journal's scope. Utilizing a river discharge dataset with an extended record length is valuable for numerous research applications. However, I found the paper to be poorly organized at times and not always clear in its presentation. Furthermore, it lacks a thorough discussion of the novel contributions of this dataset, remaining overly generic in certain aspects.

*Thank you very much for your constructive feedback. We have revised the manuscript according to your comments in order to improve the structure and clarity of the paper. We have also tried to show better the uniqueness and novelty of our dataset.*

### Abstract

The abstract is unclear and should explicitly state the manuscript's objective, which only becomes apparent after a careful reading of the paper.

*We have revised the abstract in order to make the objective of the paper clearer.*

### Introduction

The introduction includes essential material but could benefit from improvement, particularly in its concluding section, which is overly cryptic. The introduction should conclude with a clear and well-articulated statement of the manuscript's objective.

*Similar to the abstract, we have revised the introduction in order to better explain the objective of the manuscript.*

### Dataset Description and Methods

The manuscript conflates the dataset description with the methods. These should be clearly separated for better clarity. For instance:

Figure 1 is not cited in the text.

*Please check, Figure 1 is cited in the introduction (l. 58) and the methodology (l.72).*

River discharge data should separately describe the current GRDC dataset and the Caravan dataset.

*We followed the comment and added separate chapters on methodology and dataset description. The comparison between the core Caravan dataset and the GRDC extension is now improved and also highlighted in Fig. 2 and 3.*

Catchment attributes and DEM should be included as part of the dataset description.

*We think the detailed description of the methodology about the catchment attributes is sufficient and that an analysis about it would go beyond the scope of this data description paper. Since the DEM is an input data for the area calculation, the short overview of the extent and the resolution seems adequate.*

## **Outlook Section**

The outlook section lacks a crucial comparison between the datasets prior to and after the extension. It should emphasize what has been improved, the capabilities now available that were previously unattainable, and the added value of this new dataset.

*We have followed the comment and have improved the outlook section.*

## **Additional minor comments are reported in the annotated pdf:**

The abstract is not really informative of what is the aim of this paper and what is finally provided by this new dataset or this new extension. Please try to improve it.

*We have revised the abstract in order to make the aim and objective of the paper more clear.*

a new extension of what?

*The sentence has been revised. A better explanation of "extension" is now provided in paragraph 3 of the introduction.*

this sentence is not clear.

*This sentence has also been revised.*

Please explain how

*This sentence has been removed. A detailed description of how river discharge data and catchments are compiled is provided in the methodology.*

This is not very detailed. Maybe refer to the section where this is explained

*We have improved the sentence.*

This is more method than dataset description

*We have followed the reviewers comment and reorganized the manuscript into "Methodology" and "Dataset description". The figure is now part of the chapter "Methodology".*

Please clarify "base" dataset

*In the revision, we have agreed to use "core" dataset instead of "base". The definition of "core dataset" is now explained in paragraph 3 of the introduction.*

Can you explain how you handle catchments with areas lower than the grid size of era land?

*We take the forcings of the grid cell (i.e. the spatial averaged forcings of the catchment area are just defined by the forcings of that grid cell).*

*There is some degree of uncertainty, because e.g. the "rain" of a grid cell could fall entirely on the area outside of the polygon. For this very reason, the initial versions of the Caravan core dataset had a min (and also max) area defined. However, this was removed in a later version of Caravan and we follow the same principle here. With this dataset we provide a dataset to people that naturally comes with uncertainties. People using this dataset and working in environmental sciences in general, should probably be aware of the uncertainty that comes with this and depending on their research study decide which of the basins to include.*