

Interactive comment on “The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow archive and metadata” by Hong Xuan Do et al.

Anonymous Referee #2

Received and published: 31 December 2017

The paper describes the harmonised station metadata and catchment characteristics of a merged global river discharge dataset. There will be different opinions regarding the approach 'more will offset potential quality deficits' versus a 'less is more' strategy that is often applied (and necessary) for specific research. This could perhaps still be discussed a little bit better in this paper's intro. Nevertheless, I found the material well presented and the data will be useful. The steps towards the collation, selection and derivation and processing of the metadata for this large archive are well described. This documentation may help the appreciation of the often invisible but always tremendous effort that goes into harmonized datasets and I would like to highlight in particular the careful consideration and provision of quality flags for the derived metadata in this case.

C1

I hope that this information will be used, rather than overlooked. Perhaps a sentence on this important data aspect could be placed more prominently in the abstract and conclusion. Hopefully, the paper will provide incentive for some national databases, to provide access to the metadata they often have but don't provide as readily, such as catchment boundaries, topographical features and land cover.

A few minor issues that I recommend be addressed are listed below.

line 33 "questions over its utility" - it's not really clear what is meant. If the intended use is climate sensitivity analysis, yes, but there are quite a few other uses. Maybe clarify utility for... or phrase more generally.

line 324ff This section contains a bit of redundant information and a few typos (suggest to proofread again)

For the reader to get an impression of the precision of catchment area delineation, I think it is important to show an zoomed example of some kind.

In 5.2 or in the conclusion I think a bit more discussion or cautionary words should be spent over the fact that there will be catchments in the database for which streamflow time series do not overlap barely or not at all with the time covered by (the relatively new or short) remote sensing based datasets. This requires users to carefully check time overlap for possible cause-effect studies. And ideally metadata readme or column headers should provide the time period covered by the underlying datasets.

Figures 1 and 4 (upper) and 5 (upper) are entirely useless at the resolution and in the jpg format provided in the pdf-download. Dots are indistinguishable. High resolution will be necessary, but likely still not sufficient to make this a useful map. I suggest to create zooms into subdivided regions that will allow to see some of the differences within regions/countries.

Fig 5 lower. Make proper superscripts in the axes labels and change tick labels units e.g. to million or so (or at least also use proper superscripting) - see Journal's

C2

Manuscript guidelines.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2017-103>, 2017.