

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2022-9-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-9**

Anonymous Referee #2

Referee comment on "Hydrography90m: a new high-resolution global hydrographic dataset" by Giuseppe Amatulli et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-9-RC2>, 2022

This is a well written paper and nicely conducted study on extracting vector river flowlines based on the MERIT-Hydro product, for people to use in global river-related applications. However, I have the following concerns which prohibits the immediate recommendation of this paper for publication in this round.

Major concerns:

On channelization threshold: 0.05 km² does not always outline better hydrography. Because uncertainty remains in headwater channel initialization threshold, and if you read earlier HydroSHEDS documentations by Lehner's group, and you will find it says often times threshold smaller than 100 grid cells (100*90m*90m) will introduce a lot of uncertainty. So, can the authors provide more justification on why using 0.05km²? How about the uncertainty it provided? What if the 726M lines mean just "more" channels, but not "better" channels? (related to this, what does it mean by saying "we address the all-important issues of headwaters" in Line 70? Is it related to the 0.05 km² threshold?)

Data format: I have some concerns of storing the data in 20-degree by 20-degree tiles. There's no problem with grid-based MERIT DEM or Hydro products of storing the data in tiles. However, when it comes to vectors, storing the data this way will artificially cut off the channels that crosses basin boundaries. If you look at other widely used vector-based hydrography dataset (e.g. Lehner et al 2008; Linke et al. 2019; Lin et al. 2019; Lin et al 2021), they have all stored data in terms of different watershed levels. Can the authors

rethink their data supplying strategy? I think a user that accidentally sits in a basin being cut by the tiles will have a frustration using the data. Isn't it?

Figure 8: this is related to the earlier uncertainty question – how does this overlap percentage differ if considering wet/dry region differences? For example, smaller threshold is definitely better for wet regions, but how about dry regions? I think this evaluation method can exaggerate the benefits of Hydrography90m. Can the authors provide more assessments here?

Minor ones:

I disagree with the statement “no such hydrographic study has been published to date” in the Abstract. Since the publication of MERIT Hydro in 2019, there has been many hydrographic work published based on MERIT Hydro. I think the authors should step back a bit in the very beginning of this paper.

Line 51 and thereafter is a mis-citation to Lin et al. (2019). That paper used 25 km² threshold for the sake of building a model. I think a later paper by Lin et al. (2021) was the one featuring variable drainage density.

Figure 3: I appreciate this figure, but again related to my earlier comment on data supplying format – is it possible to supply data by using the basin outlined here? Otherwise readers will make additional efforts to merge the data themselves if being cut.

