Referee comment

A hydrography upscaling method for scale invariant parametrization of distributed hydrological models

Dirk Eilander, Willem van Verseveld, Dai Yamazaki, Albrecht Weerts, Hessel C. Winsemius, and Philip J. Ward

The authors describe a method to uspscale high-resolution data, based on a back-tracking approach also used by former work of Yamazaki et al. 2009 or Wu et al. 2011). The authors show methods to assess the quality of different methods.

The paper represents significant progress as it is a further development of the concept of Yamazaki et al:

- The authors can show that there approach yields better results in comparison to other methods
- It is open source
- The authors show that their method can be applied in hydrological models towards free scalable models
- It can be used as the common D8 network

The presentation quality and scientific quality is good. A few points could be discussed in a different way.

Main point of criticism is that it does not include a link to the work done in the ISI-MIP (https://www.isimip.org/) project. In this project, quite a number of hydrological models use a defined set of input data as a global 30 arcmin setting. The network used here is the DDM30 (Doell and Lehner 2002). It is questionable, if this database is the best choice (see Zhao et al. 2017 https://doi.org/10.1088/1748-9326/aa7250), but it is used as the defined river network. The paper can run without a direct comparison to DDM or a comparison to 30 arcmin, but the value of this paper (and the numbers of citations) can be improved, if it is compared against:

a.) 30 arcmin (maybe instead of 15 arcmin, which is rarely used in hydrological models)
b.) DDM (maybe instead DMM, as the DMM is not so often used (cited 25 times and DDM cited 147 times)

Using a power function of upstream area for river width is a weak point here. This does not work for a global dataset and not even for the River Rhine with high runoff in the mountains and low runoff in the lowlands. This approach is not state of the art. The paper says it will provide a parametrization for distr. hydrological models. The approach for providing river width is not appropriate. For sure it would be fine to have the full package incl. river width and Manning's roughness. I think it is still a fine paper, if you exclude river width (as you exclude Manning's anyway)

In detail comments:

60: As I said, I am not a fan of the DDM30, but it is THE reference river network in a global hydrological intercomparison project.

66: Wu et al. 2011,2012 and Yamazaki (2009) already give out length, Yamazaki already give out slope or elevation at the outlet point.

85 .. often defined by … Isn't it a requirement to be a multiple of the finer grid?

115: the equation needs $|x-x_0|^{0.5}$ instead of brackets

118: if no output pixel is found… where does this happen in fig 1

126: Maybe a description like in chess B2 instead h would be easier

154: You have several thresholds in your method e.g. sqrt(R), min upstream_area = 0.25, length of cell = 0.25. Did you test this setting, did you do a sensitive analysis of these values. Where are they from?
Maybe for the Rhine it would be good to show some variation of these thresholds

189: Nice solution of these "orphan" problem – upstream cells that have no direct parents

204-206: This paper does not show a valid way to derive river width for all cells (and it is stated well, you need the parameter for all cells). Filling up with a power function of upstream area will not work. Maybe using some regression/machine learning technique like in Barbarossa et al. 2018 (https://www.nature.com/articles/sdata201852) will help. Maybe dismiss 204-206 and 2F (and write a second paper on width and Manning's). For the routing example, your assumption of width and Manning's is ok

205: Which outlet pixel in @F does not have a river width in fig 2F. Using river width interpolation with a power function and upstream area is a really weak assumption.

222: As before: it is a dataset, but an incomplete dataset for kinematic routing missing an adequate solution for river width and channel roughness. I am not asking for these 2 additional parameters, but mentioning that these 2 are missing for a complete dataset (and maybe river depth, too).

229: Also mentioned before: DMM is rather special, a comparison to DDM30 would be better

260: Not so clear, why the minimum upstream area is chosen to be 10 km2. Is this done only for the original 3arcsec, or also for the 30 arcsec version? A 1km2 threshold would be more reasonable? Why having a network to 30 arcsec and then aggregating again to 10km2? Adding a reason for 10km2 would be ok.

265-273: For the synthetic runoff event, a simple assumption is ok, like equation 5 and roughness=0.03. What about the river depth to get the perimeter you need for the kinematic routing?

273: later in357 you describe the synthetic runoff event, It would be better to describe here in more detail how your synthetic event looks like. From your sentence in line 356 I assume: Uniform runoff for the whole Rhine of around 0.2 mm per day (0.002 per 15 min) to reach 500 m/s at run outlet and then increasing to 0.014 mm per 15 min to reach peak of 3000 m/s?

350: A table for the Rhine as the table 2 would be good. Maybe even later and including the synthetic runoff results.

371: grey line as cumulated runoff?

376-404: This part is interesting, but would benefit if it concludes in a method which can be used to compare different methods in numbers e.g. creating a table with flood peak magnitude (btw a flood at around 3000 m3/s is not a flood in the upper Rhine) and timing like the numbers in line 383 and line 385.
But due to the different N, the numbers are not really comparable. How about selecting only locations which all methods have in common, describing a method to find these locations and then it is possible to set up a table with peak magnitude smaller or bigger than 2%, 5%, 10% and flood peak timing different by percent of runoff peak time to routing peak time

384: maybe not an absolute hour, but a percentage of the difference between running time between runoff peak (gray line) and reference time (3 arcsec model). Because it does not matter so much if the delay in a big basin is 2 hours, but it matters if it is in a small basin.

442 As before: river width I would delete from this list as you cannot derive it in a proper way with IHU.

448-444: To my understanding, there is no well-established geomorphic relationship between discharge and river depth nor river width. It is mostly regression between discharge, upstream area, etc. like in Pistocci 2006 or others. A way to improve this is machine learning and feeding it with climatic, geomorphic data like Barbarossa et al. 2018 or using advanced technics of remote sensing e.g., Allen and Pavelsky 2018 or a combination of both. Your approach for having a full dataset for benchmarking is ok, but maybe dismiss the part for river width.

465: The Merit DEM (Yamazaki et al. 2017) is a very good DEM. But mostly due to anthropogenic overprinting, rivers do not always follow the lowest elevation. They are redirected or even on a higher level than the surrounding landscape (there are many examples of this in the Netherlands). One way to improve the DEM would be to burn in the existing river network. Not part of your work here, but maybe something to think of in the discussion.